

STIC-Biotech/ChemLib

93659

From: Rao, Manjunath N.  
Sent: Thursday, May 08, 2003 5:41 PM  
T: STIC-Biotech/ChemLib  
Subject: Sequence search request for 10/028,245

RECEIVED

MAY - 9 2003

STIC/BIOCHEM. DIVISION  
(STIC)

From: Manjunath N. Rao  
Art Unit 1652, Room 10A11  
Mail Box in Room 10D 01  
Phone: 306-5681

Date: 5-7-03

Please search the following as soon as possible for application with serial number **10/028,245**

1. SEQ ID NO: 1 AND 4 against all commercial nucleic acid databases including issued patents database and pending application database and provide a print of all results.
2. SEQ ID NO:2, AND 3 against all commercial protein databases including issued patents database and pending application database and provide a print of all results.

If you have any questions please call me at the above phone number.

Thanks

Edward Han  
Technical Info. Specialist  
STIC/Biotech  
CMI 6B02 Tel: 305-9203

Manjunath N. Rao, Ph.D.  
Biotechnology Patent Examiner  
Art Unit 1652, Room 10A11  
Mail Box in 10D01  
Crystal Mall, USPTO.

Searcher: \_\_\_\_\_  
Phone: \_\_\_\_\_  
Location: \_\_\_\_\_  
Date Picked Up: 5/9/03  
Date Completed: 5/14/03  
Searcher Prep/Review: \_\_\_\_\_  
Clerical: \_\_\_\_\_  
Online time: \_\_\_\_\_

TYPE OF SEARCH:  
NA Sequences: 2  
AA Sequences: 2  
Structures: \_\_\_\_\_  
Bibliographic: \_\_\_\_\_  
Litigation: \_\_\_\_\_  
Full text: \_\_\_\_\_  
Patent Family: \_\_\_\_\_  
Other: \_\_\_\_\_

VENDOR/COST (where applic.)  
STN: \_\_\_\_\_  
DIALOG: \_\_\_\_\_  
Questel/Orbit: \_\_\_\_\_  
DRLink: \_\_\_\_\_  
Lexis/Nexis: \_\_\_\_\_  
Sequence Sys.: 43/02  
WWW/Internet: \_\_\_\_\_  
Other (specify): \_\_\_\_\_

**THIS PAGE BLANK (USPTO)**



Best Local Similarity 61.1%; Pred. No. 69;  
Matches 11; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

OY 2 RATSILAAALAVAGDALA 19  
1 : | | | | | | | | | |  
Db 3 RTGALLAALALACCAQA 20

RESULT 14  
ID PLYA\_COLGL STANDARD; PRT; 380 AA;  
AC 000374;  
DT 15-DEC-1998 (Rel. 37, Created)  
DT 15-DEC-1998 (Rel. 37, Last sequence update)  
DE Pectin lyase precursor (EC 4.2.2.10).  
GN PNA.

OS Colletotrichum gloeosporioides (Anthracoaceae fungus) (Glomerella  
OS cingulata).  
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;  
OC Sordariomycetes incertae sedis; Phyllochorales; Phyllochoraceae;  
OC Glomerella.  
ON NCBI\_Taxid=5457;  
RX 1;  
RP SEQUENCE FROM N.A.  
RA MEDLINE=94237480; Pubmed=8181749;  
RA Templeton M.D., Sharrock K.R., Bowen J.K., Crowhurst R.N.,  
RA Rikkersink E.H.;  
RT "The pectin lyase-encoding gene (pnl) family from Glomerella  
RT cingulata: characterization of pnl and its expression in yeast."  
RL Gene 142141-146(1994).  
CC -1- CATALYTIC ACTIVITY: Eliminative cleavage of pectin to give  
CC oligosaccharides with terminal 4-deoxy-6-methyl-alpha-D-galact-4-  
CC enuronosyl groups.  
CC -1- SIMILARITY: BELONGS TO THE POLYSACCHARIDE LYASE FAMILY 1.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL: L22857; AAA21817.1; -  
DR HSSP: 001172; 1IDJ.  
DR InterPro: IPR002022; Amb\_allergen.  
DR Pfam: PF00544; pec\_lyase; 1.  
KW Lyase; Signal; Glycoprotein.  
FT SIGNAL 1 20  
FT CHAIN 21 380 PECTIN LYASE.  
FT CARBOHYD 130 130 N-LINKED (GLCNAC... ) (POTENTIAL).  
SQ SEQUENCE 380 AA; 39326 MW; 3DF9A99FB8482053 CRC64;

Query Match 49.4%; Score 41; DB 1; Length 380;  
Best Local Similarity 44.0%; Pred. No. 69;  
Matches 11; Conservative 5; Mismatches 3; Indels 6; Gaps 1;

OY 1 MRATSLAAALAV-----AGDALA 19  
1 : | | | | | | | | | |  
Db 1 MRSATSLAAALAPLASADAVS 25

RESULT 15  
ID YAO2\_RHIME STANDARD; PRT; 525 AA.  
AC 052969;  
DT 30-MAY-2000 (Rel. 39, Created)  
DT 30-MAY-2000 (Rel. 39, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Hypothetical sensor-like histidine kinase R01002 (EC 2.7.3.-).  
GN R01002 OR SMC00059.

OS Rhizobium meliloti (Sinorhizobium meliloti).  
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;  
OC Rhizobiaceae; Sinorhizobium.  
ON NCBI\_Taxid=382;  
RX 1;  
RP SEQUENCE FROM N.A.  
RA STRAIN=1021;  
RA MEDLINE=21396507; Pubmed=11481430;  
RA Capela D., Barloy-Hubler F., Gouzy J., Bothe G., Ampe F., Batut J.,  
RA Bolstad P., Becker A., Boutry M., Cadieu E., Dreano S., Gloux S.,  
RA Godrie T., Goffeau A., Kahn D., Kiss E., Lelaure V., Masuy D.,  
RA Pohl T., Portetalle D., Puehler A., Purnelle B., Ranspaege U.,  
RA Renard C., Thebaud P., Vandenbol M., Weidner S., Gallibert F.;  
RT "Analysis of the chromosome sequence of the legume symbiont  
RT Sinorhizobium meliloti strain 1021."  
RL Proc. Natl. Acad. Sci. U.S.A. 98:9877-9882(2001).  
RN 12;  
RP SEQUENCE OF 1-87 FROM N.A.  
RC STRAIN=RCR2011 / 5047;  
RX MEDLINE=95276304; Pubmed=7756693;  
RA Keller M., Rohlau A., Weng W.M., Schmidt M., Quandt J., Niehaus K.,  
RA Jording D., Arnold W., Puehler A.;  
RT "Molecular analysis of the Rhizobium meliloti muck gene regulating the  
RT biosynthesis of the exopolysaccharides succinoglycan and  
RT galactoglucan."  
RL Mol. Plant Microbe Interact. 8:267-277(1995).  
CC -1- SUBCELLULAR LOCATION: Integral membrane protein (Potential).  
CC -1- SIMILARITY: CONTAINS 1 HISTIDINE KINASE DOMAIN.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL: AL591785; CAC45574.1; -  
DR EMBL: L37353; AAA74242.1; -  
DR InterPro: IPR003594; ATPbind\_ATPase.  
DR InterPro: IPR004358; Bact\_sens\_pr\_C.  
DR InterPro: IPR004359; HIS\_KIN\_519.  
DR InterPro: IPR003661; HIS\_KIN.  
DR InterPro: IPR001610; PAC.  
DR InterPro: IPR000700; PAS-assoc\_C.  
DR Pfam: PF00512; signal; 1.  
DR Pfam: PF00785; PAC; 1.  
DR Pfam: PF02518; HATPase\_C; 1.  
DR PRINTS: PR00344; BCTRISENSOR.  
DR SMART: SM00387; HATPase\_C; 1.  
DR SMART: SM00388; HIS\_KIN; 1.  
DR SMART: SM00086; PAC; 1.  
DR PROSITE: PSS0109; HIS\_KIN; 1.  
DR PROSITE: PSS0113; PAC; 1.  
KW Hypothetical protein; Sensory transduction; Transferase; Kinase;  
KW Transmembrane; Phosphorylation; Complete proteome.  
FT TRANSMEM 36 56  
FT TRANSMEM 61 81  
FT DOMAIN 173 228 PAC.  
FT DOMAIN 246 466 HISTIDINE KINASE.  
FT MOD\_RES 249 249 PHOSPHORYLATION (AUO-) (BY SIMILARITY).  
SQ SEQUENCE 525 AA; 55899 MW; 53663574463F99F0 CRC64;

Query Match 49.4%; Score 41; DB 1; Length 525;  
Best Local Similarity 69.2%; Pred. No. 89;  
Matches 9; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

OY 3 ATSLAAALAVAG 15  
1 : | | | | | | | | | |  
Db 71 SASILAAALAVAG 83

Fri May 16 10:43:19 2003

Search completed: May 9, 2003, 15:18:55  
Job time : 2.73744 secs

us-10-028-245-3.rsp

Page 9

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.4.p5.4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 9, 2003, 15:15:41 ; Search time 4.64155 Seconds  
(without alignments)  
843.445 Million cell updates/sec

Title: US-10-028-245-3  
Perfect score: 83  
Sequence: 1 MRATSLAALAVAGDALA 19

Scoring table: BIOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

SPTREMBL\_21:\*

- 1: sp\_archaea:\*
- 2: sp\_bacteria:\*
- 3: sp\_fungi:\*
- 4: sp\_human:\*
- 5: sp\_invertebrate:\*
- 6: sp\_mammal:\*
- 7: sp\_mmc:\*
- 8: sp\_organelle:\*
- 9: sp\_phage:\*
- 10: sp\_plant:\*
- 11: sp\_proteus:\*
- 12: sp\_virus:\*
- 13: sp\_vertebrate:\*
- 14: sp\_unclassified:\*
- 15: sp\_virus:\*
- 16: sp\_bacteriophage:\*
- 17: sp\_archaeal:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	49	59.0	183	09A7T4	09A7T4 caulobacter
2	49	59.0	558	08S068	08S068 oryza sativ
3	46	55.4	136	092MW3	092MW3 rhizobium m
4	46	55.4	388	5 016501	016501 caenorhabdi
5	45	55.4	501	10 09SYD6	09SYD6 arabidopsis
6	45	54.2	61	10 0942E3	0942E3 oryza sativ
7	45	54.2	75	16 08U5R3	08U5R3 agrobacteri
8	45	54.2	166	16 09CGF4	09CGF4 lactococcus
9	45	54.2	171	12 071097	071097 bovine aden
10	45	54.2	178	16 09AB20	09AB20 caulobacter
11	45	54.2	317	17 08T294	08T294 methanopyru
12	45	54.2	363	16 069949	069949 streptomyc
13	45	54.2	472	16 086362	086362 mycobacteri
14	44	53.0	182	16 09KSV2	09KSV2 vibrio chol
15	44	53.0	201	4 096CH8	096CH8 homo sapien
16	44	53.0	320	16 08U676	08U676 agrobacteri

17	44	53.0	578	2 08VW87	08VW87 archibacte
18	44	53.0	604	16 092M72	092M72 rhizobium m
19	43	51.8	89	16 0821H0	0821H0 salmonella
20	43	51.8	137	10 09SLB6	09SLB6 arabidopsis
21	43	51.8	213	16 092TV0	092TV0 rhizobium m
22	43	51.8	220	2 066285	066285 agrobacteri
23	43	51.8	243	16 09R208	09R208 deinococcus
24	43	51.8	292	16 082214	082214 salmonella
25	43	51.8	304	16 0821J4	0821J4 salmonella
26	43	51.8	309	16 08R1U5	08R1U5 rhizobium l
27	43	51.8	353	16 08XYM1	08XYM1 ralsstonia s
28	43	51.8	361	16 092S90	092S90 rhizobium m
29	43	51.8	362	16 092PJ6	092PJ6 rhizobium m
30	43	51.8	368	16 09RXU5	09RXU5 deinococcus
31	43	51.8	498	16 09RX04	09RX04 deinococcus
32	42	50.6	80	10 09M0H9	09M0H9 arabidopsis
33	42	50.6	275	5 09BKH4	09BKH4 caenorhabdi
34	42	50.6	306	16 08XFP8	08XFP8 salmonella
35	42	50.6	328	2 054237	054237 streptomyc
36	42	50.6	358	2 093R90	093R90 corynebacte
37	42	50.6	388	5 016500	016500 caenorhabdi
38	42	50.6	469	10 08S6V6	08S6V6 oryza sativ
39	42	50.6	501	10 08S6Q1	08S6Q1 oryza sativ
40	42	50.6	513	4 09UFP1	09UFP1 homo sapien
41	42	50.6	888	2 085840	085840 spingomona
42	42	50.6	1125	16 09A3K9	09A3K9 caulobacter
43	41.5	50.0	129	4 08RBD4	08RBD4 homo sapien
44	41.5	50.0	329	16 098K07	098K07 rhizobium l
45	41	49.4	184	16 09EWQ4	09EWQ4 streptomyc

## ALIGNMENTS

## RESULT 1

09A7T4 PRELIMINARY; PRT; 183 AA.  
ID 09A7T4;  
AC 09A7T4;  
DT 01-JUN-2001 (TREMblrel. 17, Created)  
DT 01-JUN-2001 (TREMblrel. 17, Last sequence update)  
DT 01-MAR-2002 (TREMblrel. 20, Last annotation update)  
DE Hypothetical protein CC1636.  
DE CC1636.  
OS Caulobacter crescentus.  
OC Bacteria; Proteobacteria; alpha subdivision; Caulobacter group;  
OC Caulobacter.  
OX NCBI\_TaxID=155892;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=ATCC 19089 / CB15;  
RX MEDLINE=21173698; PubMed=11259647;  
RA Nieman W.C., Feldblum T.V., Laub M.T., Paulsen I.T., Nelson K.E., Eissen J., Heidelberg J.F., Alley M.R.K., Ohta N., Maddock J.R., Potocka I., Nelson W.C., Newton A., Stephens C., Pladde N.D., Ely B., DeBoy R.T., Dodson R.J., Durkin A.S., Gunn M.L., Haft D.H., Kolonay J.F., Smit J., Craven M.B., Khouri H., Shetty J., Berry K., Utterback T., Tran K., Wolf A., Vamathevan J., Ermolaeva M., White O., Salzberg S.L., Venter J.C., Shapiro L., Fraser C.M.;  
RA "Complete genome sequence of Caulobacter crescentus."  
RT Proc. Natl. Acad. Sci. U.S.A. 98:4136-4141(2001).  
DR EMBL: AE005839; AAK23614.1; -  
DR TIGR: CC1636; -  
KW Hypothetical protein; Complete proteome.  
SQ SEQUENCE 183 AA; 19526 MW; 98E93ED4015521EC CRC64;

## Query Match

Best local similarity 59.0%; Score 49; DB 16; Length 183;  
Matches 11; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

OY 1 MRATSLAALAVAGDALA 19  
Db 37 MFVASTIAAGLSVAGSALA 55

RESULT 2  
 ID 085068 PRELIMINARY; PRT; 558 AA.  
 AC 085068;  
 DT 01-JUN-2002 (TREMBlrel. 21, Created)  
 DT 01-JUN-2002 (TREMBlrel. 21, Last sequence update)  
 DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)  
 DE P678F11.10 protein.  
 GN P678F11.10.  
 OS Oryza sativa (Japanese cultivar-group).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
 OC Ehrhartoideae; Oryzaceae; Oryza.  
 NCBI\_TaxID:39947;  
 RN  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-CV. NIPPONBARE;  
 RA Sasaki T., Matsumoto T., Yamamoto K.;  
 RT "Oryza sativa nipponbare(GA3) genomic DNA, chromosome 1, PAC  
 RT clone: P0678F11.";  
 RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AP003437; BAB86103.1;  
 SO SEQUENCE 558 AA; 57437 MW; B1A20C9E153F4DB9 CRC64;

Query Match 59.0%; Score 49; DB 10; Length 558;  
 Best Local Similarity 75.0%; Pred. No. 42;  
 Matches 12; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

OY 3 ATSLAALAVAGDAL 18  
 DB 446 ASSLADGLAVAGAL 461

RESULT 3  
 ID 092MW3 PRELIMINARY; PRT; 136 AA.  
 AC 092MW3;  
 DT 01-DEC-2001 (TREMBlrel. 19, Created)  
 DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)  
 DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)  
 DE Hypothetical protein RB0214.  
 GN RB0214 OR SMB20222.  
 OS Rhizobium meliotti (Sinorhizobium meliotti).  
 OC Plasmid pSymb (megaplasmid 2).  
 OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;  
 OC Rhizobiaceae; Sinorhizobium.  
 NCBI\_TaxID:382;  
 RN  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-1021;  
 RA MEDLINE-21396508; Pubmed-11481431;  
 RA Flann T.M., Weidner S., Wong K., Buhrmester J., Chain P.,  
 RA Vorhoefer F.J., Hernandez-Lucas I., Becker A., Cowie A., Gouzy J.,  
 RA Golding B., Puehler A.;  
 RT "The complete sequence of the 1.683-kb pSymb megaplasmid from the N2-  
 RT fixing endosymbiont Sinorhizobium meliotti.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 98:9889-9894 (2001).  
 DR EMBL: AL603642; CAC48614.1;  
 KW Plasmid; Hypothetical protein; Complete proteome.  
 SO SEQUENCE 136 AA; 14195 MW; 3076EDD3FD67F97B CRC64;

Query Match 55.4%; Score 46; DB 16; Length 136;  
 Best Local Similarity 52.6%; Pred. No. 32;  
 Matches 10; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

OY 1 MRATSLAALAVAGDAL 19  
 DB 4 MKTSRLAALLVGSLA 22

RESULT 4  
 016501

ID 016501 PRELIMINARY; PRT; 388 AA.  
 AC 016501;  
 DT 01-JAN-1998 (TREMBlrel. 05, Created)  
 DT 01-JAN-1998 (TREMBlrel. 05, Last sequence update)  
 DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)  
 DE C03A7.7 protein.  
 GN C03A7.7.  
 OS Caenorhabditis elegans.  
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;  
 OC Rhabditidae; Pelodierinae; Caenorhabditis.  
 NCBI\_TaxID:6339;  
 RN  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-BRISTOL N2;  
 RA MEDLINE-94150718; Pubmed-7906398;  
 RA Wilson R., Alnecough R., Anderson K., Baynes C., Berts M.,  
 RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,  
 RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,  
 RA Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,  
 RA Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,  
 RA Lightning J., Lloyd C., McMurray A., Mortimore B., O'Callaghan M.,  
 RA Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Showkeen R.,  
 RA Smalton N., Smith A., Sonhammer E., Staden R., Sultson J.,  
 RA Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,  
 RA Watson A., Weinstock L., Wilkinson-Spratt J., Wohlman P.,  
 RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.  
 RT elegans.";  
 RN Nature 368:32-38 (1994).  
 RN  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-BRISTOL N2;  
 RA Greco T., Bradshaw H., Elliott G.;  
 RT "The sequence of C. elegans cosmid C03A7.";  
 RL Submitted (AUG-1997) to the EMBL/GenBank/DBJ databases.  
 RN  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-BRISTOL N2;  
 RA Waterston R.;  
 RL Submitted (JUL-1997) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AF016451; AAB65996.1;  
 DR InterPro: IPR003341; DUF139.  
 DR Pfam: PF02363; DUF139; 9.  
 SO SEQUENCE 388 AA; 42129 MW; 74ACA3953E0AF2A2 CRC64;

Query Match 55.4%; Score 46; DB 5; Length 388;  
 Best Local Similarity 59.1%; Pred. No. 82;  
 Matches 13; Conservative 2; Mismatches 3; Indels 4; Gaps 1;

OY 1 MRATSLAALAVAGDAL 18  
 DB 1 MKTSRLAALLVGSLA 22

RESULT 5  
 ID 095YD6 PRELIMINARY; PRT; 501 AA.  
 AC 095YD6;  
 DT 01-MAY-2000 (TREMBlrel. 13, Created)  
 DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)  
 DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)  
 DE F11M15.20 protein.  
 GN F11M15.20.  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.  
 NCBI\_TaxID:3702;  
 RN  
 RP SEQUENCE FROM N.A.  
 RC Federspiel N.A., Palm C.J., Conway A.B., Conn L., Hansen N.F.,  
 RA Altafi H., Araujo R., Huizar L., Rowley D., Buehler E., Dunn P.,  
 RA Gonzalez A., Kremetska I., Kim C., Lenz C., Li J., Liu S.,  
 RA Lueros S., Schwartz J., Shinn P., Toriumi M., Vysotskaia V.S.,

RA Walker M., Yu G., Ecker J., Theologis A., Davis R.W.;  
 RL Submitted (MAY-1999) to the EMBL/Genbank/DBJ databases.  
 DR EMBL: AC006085; AAD30646.1; -  
 DR InterPro: IPR002528; MATE.  
 DR InterPro: IPR001865; Ribosomal\_S2.  
 DR Pfam: PF01554; DPF0013; 2.  
 DR TIGRFAMs: TIGR00797; mate; 1.  
 DR PROSITE: PS00962; RIBOSOMAL\_S2\_1; UNKNOWN\_1.  
 SQ SEQUENCE 501 AA; 53882 MW; 21127D1C189E4BC8 CRC64;

Query Match 55.4%; Score 46; DB 10; Length 501;  
 Best Local Similarity 68.8%; Pred. No. 1e+02;  
 Matches 11; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

OY 3 ATSLAALAAVAGDAL 18  
 DB 352 ATSLADGYAVAGQAI 367

## RESULT 6

O942E3 PRELIMINARY: PRT: 61 AA.  
 AC O942E3;  
 DT 01-DEC-2001 (TREMBLrel. 19, Created)  
 DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)  
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
 DE P0480C01.23 protein.  
 GN P0480C01.23.  
 OS Oryza sativa (Rice).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
 OC Ehrhartoideae; Oryzeae; Oryza.  
 OX NCBI\_Taxid=4530;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CV. NIPPONBARE;  
 RA Sasaki T., Matsunoto T., Yamamoto K.;  
 RT "Oryza sativa nipponbare(GA3) genomic DNA, chromosome 1, PAC  
 clone: P0480C01.";  
 RL Submitted (MAR-2001) to the EMBL/Genbank/DBJ databases.  
 DR EMBL: AP003453; BAB68063.1; -  
 KW SEQUENCE 61 AA; 6636 MW; 32AE68EE03631EBE CRC64;

Query Match 54.2%; Score 45; DB 10; Length 61;  
 Best Local Similarity 66.7%; Pred. No. 21;  
 Matches 12; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

OY 2 RATSLAALAAVAGDAL 19  
 DB 42 RTTELLAARAPAPARA 59

## RESULT 7

O8U5R3 PRELIMINARY: PRT: 75 AA.  
 AC O8U5R3;  
 DT 01-JUN-2002 (TREMBLrel. 21, Created)  
 DT 01-JUN-2002 (TREMBLrel. 21, Last sequence update)  
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)  
 DE AGR\_PAT\_782P.  
 GN AGR\_PAT\_782.  
 OS Agrobacterium tumefaciens (strain C58 / ATCC 33970).  
 OC Plasmid AT.  
 OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;  
 OC Rhizobiaceae; Rhizobium.  
 OX NCBI\_Taxid=176299;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC MEDLINE=21608551; PubMed=11743194;  
 RA Goodner B., Hinkie G., Gatlung S., Miller N., Blanchard M.,  
 RA Qurollo B., Goldman B.S., Cao Y., Askenazi M., Halling C., Mullin L.,  
 RA Hummel K., Gordon J., Vaudin M., Iartchouk O., Epp A., Liu F.,  
 RA Wollam C., Allinger M., Doughy D., Scott C., Lappas C., Markelz B.,

RA Planagan C., Crowell C., Guron J., Lomo C., Sear C., Strub G.,  
 RA Cielo C., Slater S.;  
 RT "Genome sequence of the plant pathogen and biotechnology agent  
 Agrobacterium tumefaciens C58.";  
 RL Science 294:2323-2328(2001).  
 DR EMBL: AE007921; AAK90902.1; -  
 KW Plasmid.  
 SQ SEQUENCE 75 AA; 8325 MW; 06ACBE021BD055B8 CRC64;

Query Match 54.2%; Score 45; DB 16; Length 75;  
 Best Local Similarity 52.6%; Pred. No. 26;  
 Matches 10; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

OY 1 MRATSLAALAAVAGDAL 19  
 DB 1 MRAATMFDAALVAGDVA 19

## RESULT 8

O9CGF4 PRELIMINARY: PRT: 166 AA.  
 AC O9CGF4;  
 DT 01-JUN-2001 (TREMBLrel. 17, Created)  
 DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)  
 DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)  
 DE Hypothetical protein ylfF.  
 GN ylfF OR L11142.  
 OS Lactococcus lactis (subsp. lactis) (Streptococcus lactis).  
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillales;  
 OC Streptococcaceae; Lactococcus.  
 OX NCBI\_Taxid=1360;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=IL1403;  
 RX MEDLINE=21235186; PubMed=11337471;  
 RA Bolotin A., Wincker P., Manger S., Jallion O., Malarne K.,  
 RA Weisenbach J., Ehrlich S.D., Sorokin A.;  
 RT "The complete genome sequence of the lactic acid bacterium Lactococcus  
 lactis ssp. lactis IL1403.";  
 RL Genome Res. 11:731-753(2001).  
 DR EMBL: AE006346; AAK05240.1; -  
 KW Hypothetical protein; Complete proteome.  
 SQ SEQUENCE 166 AA; 19761 MW; A903B0D0227BB33B CRC64;

Query Match 54.2%; Score 45; DB 16; Length 166;  
 Best Local Similarity 56.2%; Pred. No. 53;  
 Matches 9; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

OY 1 MRATSLAALAAVAGD 16  
 DB 22 MRISLLAVALAOTSGE 37

## RESULT 9

O71097 PRELIMINARY: PRT: 171 AA.  
 AC O71097;  
 DT 01-AUG-1998 (TREMBLrel. 07, Created)  
 DT 01-AUG-1998 (TREMBLrel. 07, Last sequence update)  
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)  
 DE PVI1.  
 OS Bovine adenovirus type 3 (Mastadenovirus bo3).  
 OC Viruses; dsDNA viruses, no RNA stage; Adenoviridae; Mastadenovirus.  
 OX NCBI\_Taxid=10510;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=WB-1;  
 RX MEDLINE=98105785; PubMed=9445040;  
 RA Reddy P.S., Idamakanti N., Zakharchouk A.N., Baxi M.K., Lee J.B.,  
 RA Pyne C., Babluk L.A., Tikoo S.K.;  
 RT "Nucleotide sequence, genome organization, and transcription map of  
 bovine adenovirus type 3.";  
 RT J. Virol. 72:1394-1402(1998).

```

RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-MBR-1;
RX MEDLINE-98318755; PubMed-9654686;
RA Baxi M.K., Reddy P.S., Zakharchouk A.N., Idamakanti N., Pyne C.,
RT "Characterization of bovine adenovirus type 3 early region 2B.";
RL Virus Genes 16:113-116(1998).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-MBR-1;
RX MEDLINE-98451815; PubMed-9778793;
RA Lee J.B., Baxi M.K., Idamakanti N., Reddy P.S., Zakharchouk A.N.,
RT "Genetic organization and DNA sequence of early region 4 of bovine
RL adenovirus type 3.";
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN-MBR-1;
RX MEDLINE-99119503; PubMed-9918888;
RA Reddy P.S., Chen Y., Idamakanti N., Pyne C., Babluk L.A., Tikoo S.K.;
RT "Characterization of early region 1 and p1x of bovine adenovirus-3.";
RL Virology 253:299-308(1999).
RN [5]
RP SEQUENCE FROM N.A.
RC STRAIN-MBR-1;
RA Reddy P.S., Idamakanti N., Zakharchouk A.N., Baxi M.K., Lee J.B.,
RA Pyne C., Babluk L.A., Tikoo S.K.;
RL Submitted (OCT-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF030154; AAD09724.1;
DR InterPro: IPR004912; Adeno_VII.
DR Pfam: PF03228; Adeno_VII.1.
SQ SEQUENCE 171 AA; 18959 MW; 0EACBBI3C12519A CRC64;

Query Match 54.2%; Score 45; DB 12; Length 171;
Best Local Similarity 55.6%; Pred. No. 54;
Matches 10; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

OY 2 RATSLLAALAVAGDALA 19
DB 53 RATAALADAVATGDPVA 70

RESULT 10
O9AB20
AC O9AB20; PRELIMINARY; PRT; 178 AA.
DT 01-JUN-2001 (TREMblrel. 17, Created)
DT 01-JUN-2001 (TREMblrel. 17, Last sequence update)
DT 01-MAR-2002 (TREMblrel. 20, Last annotation update)
DE Hypothetical protein CC0414.
GN CC0414.
OS Caulobacter crescentus.
OC Bacteria; Proteobacteria; alpha subdivision; Caulobacter group.
OC Caulobacter.
OX NCBI_TaxID=155892;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-ATCC 19089 / CB15;
RX MEDLINE-21173698; PubMed-11259647;
RA Nierman W.C., Feldblyum T.V., Laub M.T., Paulsen I.T., Nelson K.E.,
RA Eisen J., Heidelberg J.F., Alley M.R.K., Ohta N., Maddock J.R.,
RA Deboy R.T., Dodson R.J., Newton A., Stephens C., Phadke N.D., Ely B.,
RA Kolonay J.F., Smit J., Craven M.B., Khouri H., Shetty J., Berry K.,
RA Uterback T., Tran K., Wolf A., Vamathevan J., Ermolaeva M., White O.,
RA Salzberg S.L., Venter J.C., Shapiro L., Fraser C.M.;
RT "Complete genome sequence of Caulobacter crescentus.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:4136-4141(2001).
DR EMBL: AE005714; AAK22401.1;
DR TIGR: CC0414;
DR InterPro: IPR000782; Bigh3_fasciclin.

```

```

DR Pfam: PF02469; Fasciclin; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 178 AA; 17830 MW; 32D35BE587E5F49 CRC64;

Query Match 54.2%; Score 45; DB 16; Length 178;
Best Local Similarity 66.7%; Pred. No. 56;
Matches 12; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

OY 2 RATSLLAALAVAGDALA 19
DB 8 RAAVLAALAAVAGPALA 25

RESULT 11
O8T294
ID O8T294; PRELIMINARY; PRT; 317 AA.
AC O8T294;
DT 01-JUN-2002 (TREMblrel. 21, Created)
DT 01-JUN-2002 (TREMblrel. 21, Last sequence update)
DE Uncharacterized protein.
GN MK0042.
OS Methanopyrus kandleri.
OC Archaea; Euryarchaeota; Methanopyri; Methanopyrales; Methanopyraceae;
OC Methanopyrus.
OX NCBI_TaxID=2320;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-AV19 / DSM 6324 / JCM 9639;
RX MEDLINE-21927647; PubMed-11930014;
RA Slesarev A.I., Mezheva K.V., Makarova K.S., Polushin N.N.,
RA Shcherbina O.V., Shakhova V.V., Belova G.I., Atavind L.,
RA Natile D.A., Rogozin I.B., Tatusev R.L., Wolf Y.I., Stetter K.O.,
RA Malykh A.G., Koonin E.V., Kozaykin S.A.;
RT "The complete genome of hyperthermophile Methanopyrus kandleri AV19
RT and monophyly of archaeal methanogens.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:4644-4649(2002).
DR EMBL: AE010305; AAM01259.1;
DR Complete proteome.
KW Complete proteome.
SQ SEQUENCE 317 AA; 33959 MW; 37DED60CB9C1B838 CRC64;

Query Match 54.2%; Score 45; DB 17; Length 317;
Best Local Similarity 57.9%; Pred. No. 95;
Matches 11; Conservative 1; Mismatches 7; Indels 0; Gaps 0;

OY 1 MRATSLAALAVAGDALA 19
DB 1 MRALALLAVLAAGGAA 19

RESULT 12
O69949
ID O69949; PRELIMINARY; PRT; 363 AA.
AC O69949;
DT 01-AUG-1998 (TREMblrel. 07, Created)
DT 01-AUG-1998 (TREMblrel. 07, Last sequence update)
DT 01-JUN-2002 (TREMblrel. 21, Last annotation update)
DE Hypothetical protein SC06574.
GN SC06574 OR SC3F9.09.
OS Streptomyces coelicolor.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycinae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1902;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-A3(2);
RA Seeger K.J., Harris D.;
RL Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-A3(2);
RA Parkhill J., Barrell B.G., Rajandream M.A.;
RL Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.

```

[3]  
 RN SEQUENCE FROM N.A.  
 RP STRAIN-A3(2);  
 RC MEDLINE-97000351; PubMed-8843436;  
 RX Redenbach M., Kleser H.M., Denapalte D., Eichner A., Cullum J.,  
 RA Khashi H., Hopwood D.A.;  
 RT "A set of ordered cosmids and a detailed genetic and physical map for  
 the 8 Mb Streptomyces coelicolor A3(2) chromosome.";  
 RL Mol. Microbiol. 21:77-96(1996).  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-A3(2) / M145;  
 RA Bentley S.D., Chater K.F., Cerdano-Tarraga A.-M., Challis G.L.,  
 RA Thomson N.R., James K.D., Harris D.E., Quail M.A., Kleser H.,  
 RA Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,  
 RA Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,  
 RA Huang C.-H., Kleser T., Larke L., Murphy L., Oliver K., O'Neill S.,  
 RA Rabinowitsch E., Rajandream M.A., Rutherford K., Rutter S.,  
 RA Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,  
 RA Warren T., Wietzorrek A., Woodward J., Barrall B.G., Parkhill J.,  
 RA Hopwood D.A.;  
 RT "Complete genome sequence of the model actinomycete Streptomyces  
 coelicolor A3(2).";  
 RL Nature 417:141-147(2002).  
 DR EMBL; AL023862; CAAL9632.1; -  
 DR InterPro; IPR000537; UBLA.  
 DR Pfam; PF01040; UBLA; 1.  
 KW Hypothetical protein.  
 SQ SEQUENCE 363 AA; 35089 MW; 7A415FB94526FBB CRC64;  
 SO  
 Query Match 54.2%; Score 45; DB 16; Length 363;  
 Best Local Similarity 84.6%; Pred. No. 1.1e+02;  
 Matches 11; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
 OY 3 ATSLAALAVAG 15  
 DB 336 ATSLAALAVAG 348  
 RESULT 13  
 ID 086362 PRELIMINARY; PRT; 472 AA.  
 AC 086362;  
 DT 01-NOV-1998 (TREMBLrel. 08, Created)  
 DT 01-NOV-1998 (TREMBLrel. 08, Last sequence update)  
 DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)  
 DE Hypothetical protein Rv0290.  
 GN Rv0290 OR MT0303.18 OR MT0303.  
 OS Mycobacterium tuberculosis.  
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
 OC Actinomycetales; Corynebacteriaceae; Mycobacteriaceae; Mycobacterium.  
 OX NCBI\_TaxID-1773;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-H37RV;  
 RX MEDLINE-98295987; PubMed-9634230;  
 RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,  
 RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekaia F.,  
 RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,  
 RA Davies R., Devlin K., Fellwell T., Gentles S., Hamlin N., Holroyd S.,  
 RA Hornsby T., Jagels K., Kirog A., McLean J., Moule S., Murphy L.,  
 RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,  
 RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,  
 RA Sultson J.E., Taylor K., Whitehead S., Barrall B.G.;  
 RT "Deciphering the biology of Mycobacterium tuberculosis from the  
 complete genome sequence.";  
 RL Nature 393:537-544(1998).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-H37RV;  
 RA Parkhill J.;  
 RL Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases;  
 RN [3]

RP SEQUENCE FROM N.A.  
 RC STRAIN-CDC 1551 / OSHKOSH;  
 RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,  
 RA Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,  
 RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,  
 RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,  
 RA Bisai W.;  
 RT "Whole genome comparison of Mycobacterium tuberculosis clinical and  
 laboratory strains.";  
 RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AL021930; CAAL7365.1; -  
 DR EMBL; AE006937; AAK44527.1; -  
 DR TIGR; MT0303; -  
 DR Tuberculist; Rv0290; -  
 KW Hypothetical protein; Complete proteome.  
 SQ SEQUENCE 472 AA; 47943 MW; 2A132D3AB0A7F3D1 CRC64;  
 SO  
 Query Match 54.2%; Score 45; DB 16; Length 472;  
 Best Local Similarity 64.7%; Pred. No. 1.4e+02;  
 Matches 11; Conservative 2; Mismatches 4; Indels 0; Gaps 0;  
 OY 3 ATSLAALAVAGDALA 19  
 DB 151 ATGVLAVAGIAGA 167  
 RESULT 14  
 ID 09KSV2 PRELIMINARY; PRT; 182 AA.  
 AC 09KSV2;  
 DT 01-OCT-2000 (TREMBLrel. 15, Created)  
 DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)  
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
 DE Hypothetical protein VC1154.  
 GN VC1154.  
 OS Vibrio cholerae.  
 OC Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.  
 OX NCBI\_TaxID-666;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-EL TOR N16961 / SEROTYPE O1;  
 RX MEDLINE-20406833; PubMed-10952301;  
 RA Heidelberg J.F., Eisen J.A., Nelson W.C., Clayton R.A., Gwinn M.L.,  
 RA Dodson R.J., Haft D.H., Hickey E.K., Peterson J.D., Umayam L.A.,  
 RA Gill S.R., Nelson K.E., Read T.D., Tettelin H., Richardson D.,  
 RA Ermolaeva M.D., Vamathevan J., Bass S., Qin H., Dragoi I., Sellers P.,  
 RA McDonald L., Utterback T., Fleischmann R.D., Nierman W.C., White O.,  
 RA Salzberg S.L., Smith H.O., Colwell R.R., Mekalanos J.J., Venter J.C.,  
 RA Fraser C.M.;  
 RT "DNA sequence of both chromosomes of the cholera pathogen Vibrio  
 cholerae.";  
 RL Nature 406:477-483(2000).  
 DR EMBL; AE004195; AAF94313.1; -  
 DR TIGR; VC1154; -  
 KW Hypothetical protein; Complete proteome.  
 SQ SEQUENCE 182 AA; 20133 MW; 0C9EE62327DA3422 CRC64;  
 SO  
 Query Match 53.0%; Score 44; DB 16; Length 182;  
 Best Local Similarity 57.9%; Pred. No. 79;  
 Matches 11; Conservative 1; Mismatches 7; Indels 0; Gaps 0;  
 OY 1 MRATSLAALAVAGDALA 19  
 DB 1 MKRTIVATALLVAGSALA 19  
 RESULT 15  
 ID 096CH8 PRELIMINARY; PRT; 201 AA.  
 AC 096CH8;  
 DT 01-DEC-2001 (TREMBLrel. 19, Created)  
 DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)  
 DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)  
 RN [3]

DE Hypothetical 22.1 kDa protein.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=COLON;  
 RA Strausberg R.;  
 RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; BC014228; AAH14228.1;  
 DR InterPro: IPR001107; Band\_7.  
 DR Pfam: PF01145; Band\_7; 1.  
 KM Hypothetical protein.  
 SO SEQUENCE 201 AA; 22097 MW; E8EDD01562BD1D63 CRC64;

Query Match 53.0%; Score 44; DB 4; Length 201;  
 Best Local Similarity 52.9%; Pred. No. 87;  
 Matches 9; Conservative 3; Mismatches 5; Indels 0; Gaps 0;  
 QY 2 RATSLLAALAVAGDAL 18  
 :|||: |||||  
 Db 171 KATELIIVSLATAGDGL 187

Search completed: May 9, 2003, 15:20:52  
 Job time : 8.6415 secs

GenCore version 5.1.4.p5\_4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

## OM nucleic - nucleic search, using sw model

Run on: May 15, 2003, 15:36:50 ; Search time 3402.49 Seconds

(without alignments)  
11264.793 Million cell updates/sec

Title: US-10-028-245-4

Perfect score: 1317

Sequence: 1 atgcgcgaacctccctctc.....acatgctgtagcgtttga 1317

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapept 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Database :

GenEmbl:\*

- 1: gb\_ba:\*
- 2: gb\_hlg:\*
- 3: gb\_in:\*
- 4: gb\_om:\*
- 5: gb\_ov:\*
- 6: gb\_pal:\*
- 7: gb\_ph:\*
- 8: gb\_pl:\*
- 9: gb\_pr:\*
- 10: gb\_ro:\*
- 11: gb\_sts:\*
- 12: gb\_sy:\*
- 13: gb\_un:\*
- 14: gb\_vl:\*
- 15: em\_ba:\*
- 16: em\_fun:\*
- 17: em\_hum:\*
- 18: em\_in:\*
- 19: em\_mu:\*
- 20: em\_om:\*
- 21: em\_or:\*
- 22: em\_ov:\*
- 23: em\_pat:\*
- 24: em\_ph:\*
- 25: em\_pl:\*
- 26: em\_ro:\*
- 27: em\_sts:\*
- 28: em\_un:\*
- 29: em\_vl:\*
- 30: em\_hlg\_hum:\*
- 31: em\_hlg\_inv:\*
- 32: em\_hlg\_other:\*
- 33: em\_hlg\_mus:\*
- 34: em\_hlg\_pin:\*
- 35: em\_hlg\_rod:\*
- 36: em\_hlg\_mam:\*
- 37: em\_hlg\_vrt:\*
- 38: em\_sy:\*
- 39: em\_hlgo\_hum:\*
- 40: em\_hlgo\_mus:\*
- 41: em\_hlgo\_other:\*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	298.8	22.7	1373	8	PJEG12
2	289.8	22.0	1720	8	AB021657
3	286.6	21.8	1365	8	MP013914
4	275.6	20.9	1155	6	AR030401
5	275.6	20.9	1849	6	TREGLIT
6	150	11.4	1372	6	E37750
7	133.6	10.1	12092	1	AE012096
8	128	9.7	11366	1	AE011626
9	102.6	7.8	1540	1	PSEGL
10	101	7.7	216050	1	AL646076
11	92.2	7.0	1185	8	AF440003
12	89.8	6.8	2010	1	AF233448
13	80.4	6.1	9897	1	AE004077
14	79.8	6.1	1240	8	AF054512
15	79	6.0	1008	6	AX254752
16	78	5.9	1775	8	CPCCMC1
17	78	5.9	912	6	S45137
18	77.8	5.9	1170	8	MP014948
19	77.6	5.9	1170	8	AF331518
20	76.2	5.8	1303	8	AF295275
21	72.4	5.5	694	1	AF295276
22	72.4	5.5	694	1	AF295277
23	72.4	5.5	694	1	AF295278
24	72.4	5.5	694	1	AE011625
25	72	5.5	10597	1	AE295254
26	71.2	5.4	688	1	AF295250
27	70.8	5.4	688	1	AF295251
28	70.8	5.4	688	1	AF295252
29	70.8	5.4	688	1	AF295253
30	70.8	5.4	688	1	AF295254
31	70.8	5.4	688	1	AF295270
32	70.8	5.4	694	1	AF295271
33	70.8	5.4	694	1	AF295272
34	70.8	5.4	694	1	AF295273
35	69.2	5.3	694	1	AF295274
36	69.2	5.3	694	1	AF295261
37	69.2	5.3	694	1	AF295264
38	68.6	5.2	694	1	AF295266
39	68.6	5.2	694	1	AF295267
40	68.6	5.2	694	1	AF295268
41	68.6	5.2	694	1	CNS01AOP
42	68.6	5.2	720	8	AF295257
43	68.6	5.1	694	1	AF295258
44	67	5.1	694	1	AF295258
45	67	5.1	694	1	AF295258

## ALIGNMENTS

RESULT 1	LOCUS	DEFINITION	ACCESSION	VERSION	KEYWORDS	SOURCE	ORGANISM	REFERENCE	AUTHORS	TITLE
PJEG12	PJEG12	P. janthinellum mRNA for endoglucanase2.	X89564	X89564.1	GI:984165	eg12 gene; endo-1,4-beta-glucanase; endoglucanase 2.	Penicillium janthinellum.	1 (bases 1 to 1373)	Merrittz, G., Koch, A., Henrissat, B. and Schulz, G.	Endoglucanase of Penicillium janthinellum - sequence, heterologous expression and promoter analysis

JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 1373)  
AUTHORS Mernitz,G.  
TITLE Direct Submission  
JOURNAL Submitted (12-JUL-1995) G. Mernitz, Institut fuer Mikrobiologie,  
Max-Reimann-Strasse 16, D 14532 Kleinmachnow, FRG  
REFERENCE 3 (bases 1 to 1373)  
AUTHORS Mernitz,G., Koch,A., Henrisat,B. and Schulz,G.  
TITLE Endoglucanase II (EII) of Penicillium janthinellum: cdna sequence,  
heterologous expression and promoter analysis  
JOURNAL Curr. Genet. 29 (5), 490-495 (1996)  
MEDLINE 96207475  
PUBMED 8625430

FEATURES  
source Location/Qualifiers  
1..1373  
/organism="Penicillium janthinellum"  
/strain="Blourge IMET 43733 (C41)"  
/db\_xref="taxon:5079"  
/chromosome="3"  
/clone.lib="Lambda ZAPII"  
1..1233  
/gene="egl2"  
1.:1233  
/gene="egl2"  
/EC\_number="3.2.1.4"  
/codon\_start=1  
/product="endoglucanase 2"  
/protein\_id="CAA61740.1"  
/db\_xref="GI:984166"  
/db\_xref="SPRMBL:Q12665"  
/translation="MMAODKRTLLALLALGSLAGOOTAMCGCGGCMGTATTCVGY  
YCSQNNMYSGCLPGAASSTSTTTTTHVSSTLTKTTSSTSAPTGKAFAGVNLA  
GFGEVVNTSCGLDLOVDVEDVCNMSHENVADFNIRLPISAPGVNNNGSL  
DSNFKKEQVGSLSLSGACICYDIHNRYANRGVYGOGPPDDOTISLTQLATRY  
KSNRVFEGINPEPHDLATFWAATVKRYVALRNCAITSDMTILLPETDYTSANFTE  
NGSAAAILPVNPDPGSTNLFLIEVHKLYLSDNSGTHAEVTYNADARNLATWLRSKK  
RQALSETGGGNVSCATYMCQQDLIDIANSDVLYGWTSMNAGGFQASMNVLTEVPV  
NGVQYLIVQCFYVRKKMN"

BASE COUNT 322 a 382 c 341 g 328 t  
ORIGIN

Query Match 22.7%; Score 298.8; DB 8; Length 1373;  
Best Local Similarity 58.8%; Pred. No. 5.6e-45;  
Matches 516; Conservative 0; Mismatches 362; Indels 0; Gaps 0;

Dy 86 TTCGCCGAATGCACCTTGGCTGCACATCGACGGCAGCTGTCGCCACTACACAGTCGTG 145  
Db 287 TTGCTGGTGTCACAATTCGCCGCTTCGACTTGGCGTGTGCACAAGTGTGATCTACGACC 346  
Oy 146 TGCCCCCTGCTGAGCACAAAGAGAGATGGCGCCGCGCAGATGCAAGCATTTCCGCCAAG 205  
Db 347 TGTCTCAGTGTGATGATGATGATCGGTGACGGGTGTATAACCAAGTGTCAACATTCTGTAACG 406  
Oy 206 ACGAGCGCTCAACGCTCTTGGCATATCCGCTACATGACAGTTTGTCTTCAACAACACGG 265  
Db 407 CGGATACCTTCAACATTTCGCCCTGCGCAGCTGAGTAGTCAGTTCTTCAACAACAAC 466  
Oy 266 TGGAGCGCAAGTGGACAGCTCAACCTGGGGCTCTCTAACACAAGTGTCTCAACGCTGTC 325  
Db 467 TTGGGGGTTCTCTCCACGAAACCTTGGCAAGTATGAACAAAGTTGTTCAGGGCTGT 526  
Oy 326 TCGAAGAGGGGCGCTACTGCATGATTTGATGATGACAACTTGGCCGCGACAAAGCGGCA 385  
Db 527 CTCCTCTGCGGCTTATTTGTTGTTATTCACAACTACGCCCGCTGGAAAGCGGTGGCG 586  
Oy 386 TCATCGGCGCAGGAGCGGTGTGCGAGCAATCTTTGTGCACTCTGGGTCCAGATTCGCA 445  
Db 587 TTATCGGACAAAGCGCGTCCAACCGATGACCAATTATCATGAGCTGTGAGCAGAATTGGCA 646  
Oy 446 AGTACTAGAGACACAGCAAGAATCATCTTTGGCTATATAGAGAGCCGACGACTCG 505  
Db 647 CTCTACTAAGAGTAAAGCAGAGGTTACTCTTTTTTTTGGATCATACAGAGCCCATGATCTCA 706

OY	506	ACATTGACATCTGGGGCGGACAGCTGCCAAAGAGTGTGTACATGCCGATCCGAAAGGCCGGCG	565
Db	707	ACATGGCAACCTGGGCTGGCTCCACAGTCCAGAAGACTGTGATCGCATGCAGTAACAACCTGGAG	766
OY	566	CCACCTCGCAGATGATGATTCCTCGCCCGGACCAACTTGTGGCAGCGCTCCAGACGTATGTGT	625
Db	767	CCACGCTTCAGATGATTTGTGGTCCCTGGTACGAGATACACCAGCGCTCCAACTTCATTTG	826
OY	626	CCACTGCGCAGCGGGAAGCCCTCGGCAAGATTACGAACCCGGATGGGAAGCACCGATTGGC	685
Db	827	AGAAATGGCTCCGGTGCACAGCGCTCTCTACCGGGTGACCAACCCGGATGGATGCACACAGAAC	886
OY	686	TGTACTTTGATGTGCACAGAATATCTGCATATCAACATCCGGGGTGGCAGCCGAGTSCA	745
Db	887	TGATCTTTGATGTTCACAAGTACTGTGATAGTGAACAAAGGGGACCCACGCCGAGTSCG	946
OY	746	CCACAGACACAGCTGCAGCCCTTCAACGACTTGGGGGACTGGCTGTGAGGCACAAAGCGCC	805
Db	947	TGACCAACCAATGGCCAGCCCTTCAACCAATCTTGGCAACCTGGCTGGGCTCAAAAGAGAGAC	1006
OY	806	AGGCCATCATCTCCGAAACGGGGCGGTGCATGGAAACCTTCGAGCATGTCTGCTCGG	865
Db	1007	AGGCTCTGCTTTCCGAGACTGTGTGGCGGCAACGTCACAAAGTTGTGCCACTTACATGTGCC	1066
OY	866	CCCAGAACAGGGCCATTAGCGAAAAACAGCGAGCTGTACATTGGCTTTGTGGGCTGGGGTG	925
Db	1067	AGCAGCTTGATATCTCTCATATGCCAAATTTCTGATGTCTACCTGGCGGACCTTGGAGCG	1126
OY	926	CCGGCAGCTTTGACACGTCGTACATCTTGACTGACTGACT	963
Db	1127	CTGGAGGCTTCACAGGCTCTTTGGAATCTGTCTCTGACT	1164

FEATURES	source
LOCUS	AB021657
DEFINITION	AB021657 1720 bp DNA linear PLN 06-FEB-1999
ACCESSION	AB021657
VERSION	AB021657.1 GI:4062992
KEYWORDS	endoglucanase II.
SOURCE	Trichoderma viride (isolate:MC300-1) DNA.
ORGANISM	Trichoderma viride Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes; Hypocreales; mitosporic Hypocreales; Trichoderma. 1 (bases 1 to 1720)
REFERENCE	Watananbe,M. Endoglucanase II-T. viride Published Only in Database (1998) 2 (bases 1 to 1720)
AUTHORS	Watananbe,M.
TITLE	Direct Submission
JOURNAL	Submitted (20-DEC-1998) Manabu Watananbe, Pharmaceutical Technology Laboratories, Meiji Seika Kaisha, Ltd., Kayama 788, Odawara-shi, Kanagawa 250-0852, Japan (E-mail:watanab@mc.mesh.ne.jp, Tel:81-465-37-5106, Fax:81-465-36-2888)
AUTHORS	
LOCATION	Location/Ouallifiers 1. .1720 /organism="Trichoderma viride" /isolate="MC300-1" /db_xref="taxon:5547" join(171..499,683..1610) /gene="eg12" join(171..499,683..1610) /gene="eg12" /codon_start=1 /product="endoglucanase II" /protein_id="BAA36216.1" /db_xref="GI:4062992" /translation="MNRTNAPLLLAASILPFGAAAOOTVWGCGCGIGMSGPTSCAPGS ACSTNAYVACIIPGASTITSPRPSPGPTTRATSTSSPPSSGVAFACVNIAG FDGCGTIDGTCVTSKYVPLPKNTGANNPDGICQMOHYNDGNTIFRLPVGWQLY NNNGIGLDTSTISKYDQLVQGLSLGVCYI IDIHNYARNNGGI IGQGGPTNAQFTSL WSOLASTYASQSSRWFCIGINNEPDPVNIIMAAVYVAEYVTAIRNAGATSOYISLPGNDY



```

QY 363 CTTTCCCGCTACACAGCGCGCATATCGCCAGGAGCGGTGTGGACGACATCTTTGT 422
    || |||||
Db 514 CTACCCCTGTGTGAAAGCGCCAGATCATCGCCAGGTGGCCCAACCAAGACAGTCTGT 573
QY 423 CGACCTCTGGGTCCGATGCGAAAGTACAGAGGACAGCAAGATCATCTTTGGCT 482
    || |||||
Db 574 CTCGCTGTGGAAGCGCTGTGCTACAAAGTACAAAGGACCAACAGTGGCGCTTGGGGT 633
QY 483 GATGAAGACGCGCAGCAGC---CTCGACATTGAGTCTGGGGGCGAGCGTGGCAAAAGT 539
    || |||||
Db 634 CATGAATGAGCCCGCAGCAGTGGCCGACATCAACAAAGTGGCGGACAGGTTCACAGGT 693
QY 540 CGTCACTGCGATCCGAAAGGCGCGGCACTTCGACATGATCTCTGCGCCGGAACCA 599
    || |||||
Db 694 CGTCAACCCGATCCGCAATCGTGGCTACGACGCAATACGTCCTCCGCGGCAATGA 753
QY 600 CTTTCCGAGCGTCCGAGGATGTGTCCACTGGGCGAGCGCGGAGCGCCCTCGCAATATTC 659
    || |||||
Db 754 CTGGACGAGCGCTGGCGCTTCATTGATTAATGGCTCTCGCGCGCGTGAAGAAAGTGC 813
QY 660 GAACCCGATGGAAGACACCGATTTGCTGTACTTTGATGCCAGATATCTCGACATCA 719
    || |||||
Db 814 CAACCCGAGCGCTGACGACGCAATCTGATCTTGACGTGCAAGTACCTGACAGCGA 873
QY 720 CAATCCCGGTGCGACGCGGAGTGCACACAGACAGCTGCAC--GCTTCACAGACTT 776
    || |||||
Db 874 TAACCTGGGCGACGCACTACTGAGTGCCTGACAAACACATCGACATGCTCAACCGCT 933
QY 777 CGCGGAGTGGGTGAGGAGGAGAAACAGCGGCGCATTCATCTCCGAACGCGCGCTCAT 836
    || |||||
Db 934 TGCCGACTGTGTGAGGAGAAACAGCGCATGCGCATCAACGCGGAGCGGCGGCA 993
QY 837 GGAACCTTCGATGACTGCTTCTGCGCCAGAACAGAGCCATTAACGAAACAGCA 896
    || |||||
Db 994 CACGACACTTGTGCGAAGAAATCTTGTGAGAGAGATTCAGTACTGACAAATAGCGCA 1053
QY 897 CGTCACTTTGGCTTTTGGGTGGGTGCGGAGCTTTGACACAGTGTACATCTTAC 956
    || |||||
Db 1054 TGTCTTCTCGCTTACACTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1113
QY 957 TCTGACTCGCGCGGAGAGCC--GGGCAACTAACCCGCAACAGCTCATGACAGAGC 1014
    || |||||
Db 1114 CCAGACGCGGCAAGCACTGCGATGGCAGTACAAAGATCTGCGCGAGAGAGTC 1173
QY 1015 ATTCTGG 1021
    || |||
Db 1174 ACTGTG 1180

```

```

RESULT 4
AR030401      1155 bp   DNA      Linear   PAT 29-SEP-1999
LOCUS         AR030401
DEFINITION    Sequence 15 from patent US 5861271.
ACCESSION     AR030401
VERSION       AR030401.1 GI:5943615
KEYWORDS
SOURCE        Unknown.
ORGANISM      Unknown.

```

```

REFERENCE 1 (bases 1 to 1155)
AUTHORS    Fowler,T., Clarkson,K.A., Ward,M., Collier,K.D. and Larenas,E.
TITLE      Cellulase enzymes and systems for their expressions
JOURNAL    Patent: US 5861271-A 15 JAN-1999.
FEATURES
SOURCE     1..1155
           /organism="Unknown"

```

```

BASE COUNT    272 a      302 c      300 g      281 t
ORIGIN
Query Match    20.9%; Score 275.6; DB 6; Length 1155;
Best Local Similarity 59.2%; Pred. No. 1e-40;
Matches 489; Conservative 0; Mismatches 334; Indels 3; Gaps 1;

```

```

QY 172 GATGGCGCCGCGCAGATGAGCATTTGCGCGAAGACGAGCGGCTCAAGCTTTGCTA 231
    || |||||
Db 298 GATGGCATCGCGCAGATGACAGACTTGTCAACGAGAGAGCGGATGATATTTCCGCTTA 357
QY 232 TCCGCTACATGGCATTTGTCTCTCAACAACAGTGGAGCGGCAAGCTGGAGAGCTCAAC 291
    || |||||
Db 358 CCGTGTGAGTGGCACTTACCTGCTCAACAACAATTTGGGCGGCAATCTTGATTCACGAGC 417
QY 292 TGGGCTCTCTCAACAAGGCTGTCACAGCGCTGTCTGAGAGCGGCGGCTACTGATGAT 351
    || |||||
Db 418 ATTTCCAGTATGATGACACTGTTGTAGGGGTGCTGTCTGTGGGCGCATGCTGCTG 477
QY 352 GACATGACCACTTTGCGCGCTACACGCGCGGACATCATCGCGGAGCGGCTGTGAGC 411
    || |||||
Db 478 GACATCAACAATTAATGCTGATGAGAACGCGTGGATCATTTGGTCAAGCGCGCGCTACTANT 537
QY 412 GACATCTTTGTCGACCTGTGGTCTCAGATTCGCAAGTACTACAGAGACACAGCAAGTTC 471
    || |||||
Db 538 GCTCAATTCAGAGCGCTTTGGTGTGCGAGTGGCATCAAGTACGATCTCAGTTCAGTGGGTG 597
QY 472 ATCTTTGGCTGATGAGAGCGCGGACGACCTTCGACATTTGATGTGGGCGGAGACGTCG 531
    || |||||
Db 598 TGTTCGCGATCATGATGATGAGCGCCGACGACGTGAACATCAACACTGGGCTGCCAGCTG 657
QY 532 CAAAAGTCTGTCATGCGATCGGAAGGCGCGCGGACCTGCGAGATGATCTCTCTGCC 591
    || |||||
Db 658 CAAGAGTGTGTAACCGCAATCGGAAGCGTGGTGTGCTGACGTGCAATCATCTCTTGGCT 717
QY 592 GGAACCAACTTTGCCAGCGTGGAGACGTATGTCTCACTGGCAGCGCGGAGACCTCGCG 651
    || |||||
Db 718 GGAATGATTTGGCAATCTGTGGGCTTTCATATCCGATGAGTGCAGTGCAGCGCCCTGTCT 777
QY 652 AAGATTACGAACCGGATGAGAGCAGCATTTGCTGATTTGATGTGCCAAGATATCTC 711
    || |||||
Db 778 CAAGTACGAACCGGATGGGTCAACAACGATCTGATTTTGAAGTCAACAATATCTG 837
QY 712 GACATCAACAACCTCGGCTGCGACGCGGAGTGCACACAGCAGCAGC--GTGAGCGCTTC 768
    || |||||
Db 838 GACTGAGCAACACTCGGATGACGACGCGGATGTACTACAAATTAATGAGCGGCGCTT 897
QY 769 AACGACTTGGGAGTGGCTGAGGAGCAACAAGCGCGGCGGCTATCTCCGAACAGGCG 828
    || |||||
Db 898 TCTCGCTTGCACCTTGTGCTCGACGAACAATGCGGAGGTATCTCTACGAACACCGGT 957
QY 829 GCGTTCATGGAACCTTGTGATGATGCTCTTGTGCGCGGCGGAGCAAGCGCATTAAGCAA 888
    || |||||
Db 958 GGTGGCAACGTTCACTGCTGATACAGACATGTGCCAGCAATTCATCTCAACGAG 1017
QY 889 AACAGCGAGCTCTACATTTGCTTGTGGGTGGGTGCGGCGGAGCTTTGACACGTCGAC 948
    || |||||
Db 1018 AACTCAGATGCTTAATCTTGTGCTATGTTGGGTGCGGAGTATTTGATGAGACGAT 1077
QY 949 ATCTTGATCTGACTCCCTCGGCAAGCGCGGCACTACACGCA 994
    || |||||
Db 1078 GTCTGACGGAACACGAGTACGAGTGAATCACTCATGAGGACGA 1123

```

```

RESULT 5
TRREGIII
LOCUS         TRREGIII      1849 bp   DNA      Linear   PLN 05-MAY-1994
DEFINITION    T. reesei endoglucanase. III gene, complete cds.
ACCESSION     M19373
VERSION       M19373.1 GI:170548
KEYWORDS      endoglucanase.
SOURCE        T. reesei (strain VTT-D-80133) cDNA to mRNA, and DNA.
ORGANISM      Hypocrea jecorina
              Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
              Hypocreales; Hypocreaceae; Hypocrea.

```

```

REFERENCE 1 (bases 1 to 1849)
AUTHORS      Saloheimo,M., Lehtovaara,P., Penttila,M., Teeri,T.T., Stahlberg,J.,
              Johansson,G., Pettersson,G., Claessens,M., Tonne,P. and
              Knowles,J.K.

```







Db 2293 GGACCCCGGCAACATCTGGCTTCGAGCGCATCAATCTCTGACGCGCTACAGGCG 2234  
 QY 729 GTCCGACGCGAGTGTGACAGACAGCAACAGCTGACGCTTCA-----CGACTTCGCGGA 782  
 Db 2233 CACCAAGCCGCAATGTGACAGCTGACGCTGCGGCGAGACAGTGTGCGGCTTACCGA 2174  
 QY 783 CTGGCTGAGCAGAACAGCGCCAGGCGCATCTCCGAAAGCGCGCTCATGGAAC 842  
 Db 2173 CTGGCTGCGGCAAGCAAGAGGCGTCTGGCGAGTTCGGCAGCGCCCAACATGC 2114  
 QY 843 TTGCGTACTGCTCTCTGCGCCAGAACAGGCGCATGAGCGAAACAGCGAGCTCA 902  
 Db 2113 GGTCTGACAGCAGCGCTGAAGAGATCTGACTACATGGAAGAACGACGAGCTGTG 2054  
 QY 903 CATTGGCTTTTGGGCTGGGCTGGCGAGCTTTGACACGCTGATCATCTTGACTGAC 962  
 Db 2053 GCTGGCGTGAACCTGTGTGGGCGCGCGGCGGTGTGAGAGCGGAGCTACCGCTTCAATGT 1994  
 QY 963 TCCCTCGCGCAAGCCCGCAACTACA 988  
 Db 1993 GCAGCGGCGCAAGCAGCGCAGCA 1968

RESULT 8  
 AE011626/c 11366 bp DNA linear BCT 29-MAY-2002  
 LOCUS Xanthomonas axonopodis pv. citri str. 306, section 4 of 469 of the  
 DEFINITION complete genome.  
 ACCESSION AE011626 AE008923  
 VERSION AE011626.1 GI:21106069  
 KEYWORDS  
 SOURCE  
 ORGANISM  
 Xanthomonas axonopodis pv. citri str. 306.  
 Xanthomonas axonopodis pv. citri str. 306.  
 Bacteria; Proteobacteria; gamma subdivision; Xanthomonas group;  
 Xanthomonas.

REFERENCE  
 AUTHORS  
 1 (bases 1 to 11366)  
 da Silva, A.C.R., Ferro, J.A., Reinach, F.C., Farah, C.S., Furlan, L.R.,  
 Quaggio, R.B., Monteiro-Vitorello, C.B., Van Sluys, M.A., Almeida  
 Jr., N.F., Alves, L.M.C., do Amaral, A.M., Bertolini, M.C.,  
 Camargo, L.E.A., Camarotte, G., Cannavan, F., Cardoso, J.,  
 Churino-Santos, J.R., El-Dorry, H., Faria, J.B., Ferreira, A.J.S.,  
 Ferreira, R.C.C., Ferro, M.I.T., Formighieri, E.F., Franco, M.C.,  
 Greggio, C.C., Gruber, A., Katsuyama, A.M., Kishi, L.T., Leite  
 Jr., R.P., Lemos, E.G.M., Lemos, M.V.F., Locali, E.C., Machado, M.A.,  
 Medeiros, A.M.B.N., Martinez-Rossi, N.M., Martins, E.C., Meidanis, J.,  
 Menck, C.F.M., Miyaki, C.Y., Moon, D.H., Moreira, L.M., Novo, M.T.M.,  
 Okura, V.K., Oliveira, M.C., Oliveira, V.R., Pereira, J.H.A.,  
 Rossi, A., Sena, J.A.D., Silva, C., de Souza, R.F., Spinola, L.A.F.,  
 Takita, M.A., Tamura, R.E., Teixeira, E.C., Tezza, R.I.D., Trindade dos  
 Santos, M., Truffi, D., Tsai, S.M., White, F.F., Setubal, J.C. and  
 Kitajima, J.P.  
 Comparison of the genomes of two Xanthomonas pathogens with  
 differing host specificities  
 Nature 417 (6887), 459-463 (2002)  
 22022145  
 12024217

TITLE  
 JOURNAL  
 MEDLINE  
 PUBMED  
 REFERENCE  
 AUTHORS

FEATURES  
 source  
 TITLE  
 JOURNAL  
 Direct Submission  
 Submitted (28-NOV-2001) Departamento de Biolumina, Universidade de  
 Sao Paulo, Av. Prof. Lineu Prestes 748, Sao Paulo, SP 05508-900,  
 Brazil

Location/Qualifiers  
 1. 11366  
 /organism="Xanthomonas axonopodis pv. citri str. 306"  
 /strain="306"  
 /db\_xref="taxon:190486"  
 /note="pathovar: citri"  
 complement(349..1401)  
 /gene="egl"  
 /note="XAC0029"  
 complement(349..1401)  
 /gene="egl"  
 /note="Identified by sequence similarity; putative; ORF  
 located using Blastx/Glimmer/genemark"  
 /codon\_start=1  
 /transl\_table=11  
 /product="cellulase"  
 /protein\_id="AA034921.1"  
 /db\_xref="GI:21106070"  
 /translation="MKSLLTHTRRLRCRAISFLILAMPILAQAQSAATGLKYAGVNSLG  
 AETKSSQPCVNLNDYRPADEAYFAGKHMTNVRPLILMERLOPAGSELPAQILA  
 LIRQAVANAKAAKAYIVDVHNAVYGHGHTIGSTKVPISFENDLRLALAFSDNAV  
 IFGLMNEPYIDPOAMATPAQASIDITRKGTAKLILVPGALWTGASNSYTAQGSN  
 AVAMASIRDPILNRFAIEAHQYLDADSSGSGGCVSATVGEYLRSEFTMLRKKRF  
 LAERGTGNTVTCNIALNGMLCYMNSNSDVMWGMNMAAGMNSGATFYFNQPDAGQND  
 KPNQSIISPARRTD"  
 complement(2179..3252)  
 /gene="egl"  
 /note="XAC0030"  
 complement(2179..3252)  
 /gene="egl"  
 /note="Identified by sequence similarity; putative; ORF  
 located using Blastx/Glimmer/genemark"  
 /codon\_start=1  
 /transl\_table=11  
 /product="cellulase"  
 /protein\_id="AA034922.1"  
 /db\_xref="GI:21106071"  
 /translation="MERTPPYFPRRLAIALALILLAAQPLAQAQKEPPRPBETAGKY  
 LVDNALSCLEIKRSRKKPGLTDFKDYTPAASDESPAGGNTTIRPLMEVQPELNP  
 LGQALQILIRKSELEAKANKQYLLIDHNATYSFGKRGITSDVPAGLADIMRLALE  
 FKDKKAVIFGLMNEPNCISAPDANNAOQITTAIRTKGAKNLILVPGATAYTGASNS  
 TSYGSNAKALELIDKPNLAPFAHQYLDKDYSGTPTSDSGGEKLGFTSMR  
 ENKQKGLGEFRATANNVPCRALEGLITVEKNSDVLCGTTWAAGAKMKPDPPTYQ  
 PGKQSDKPPQALILSKYARRAK"  
 3560..4618  
 /gene="yank"  
 /note="XAC0031"  
 3560..4618  
 /gene="yank"  
 /note="Identified by sequence similarity; putative; ORF  
 located using Blastx/Glimmer/genemark"  
 /codon\_start=1  
 /transl\_table=11  
 /product="alcohol dehydrogenase"  
 /protein\_id="AA034923.1"  
 /db\_xref="GI:21106072"  
 /translation="MSTAAHYAAQAADQPLAPFERRAPGNDVOIDAYCGVCHSD  
 LHTAAHNMHTLYPSVGHIEIVGRVAVGAAVGVFGYGLDAGVCSWDSRSCASCOE  
 GEQVDCAGGTGYNGMFGGEGTYGSDHLYVDQKYLRTIHSQNLAAVAPLCA  
 GITTYEAKWKKVGPQKQVGVGLGIGHAAVYKIAKMGATVTLFTTSESKRADALRT  
 GASEVYISKDEAQAQNTLDITLNTVAPAPNHLDPFLNKLKDKGAVLYVGEQSHP  
 SPAPFNLVWKRRRLASLIGGINTQDEMDFCAKHNTVSDIETRADQINEAVERMLK  
 SDVYRFLVIDMARDQA"  
 complement(5153..6634)  
 /gene="gld"  
 /note="XAC0032"  
 complement(5153..6634)

/gene="g1dp"  
 /note="identified by sequence similarity: putative, ORF  
 located using Blastx/Glimmer/Genemark."  
 /codon\_start=1  
 /transl\_table=11  
 /product="glutamate synthase beta subunit"  
 /protein\_id="AAM34924.1"  
 /db\_xref="GI:21106073"  
 /translation="MSRQAQFQFLDLPRTMETRIPVELRTSGDMGELGRPKDAQY  
 OAGRLDLCGNCPEYCKWKCVPVNAHPOMLOLVQMBEDRIIEAALACHSNPLPEVCGVAC  
 DRILGSCSTLEEFQAVTIGAVEKYIVDAFMAGMPTDIGNYVAGSHRAYVAGVAPAGI  
 ACADILTAAGIEAIVYDREIQIGLGLGICISPKDKSVIKREITEGMCVORGLN  
 EGRVSYIEQLGAEVDAYELGATRYVDGSLPGDKLVNIPRLTEIYONSRIYRGLN  
 PGHRIADMGEDLADPLLTGKRVVYITGAGVAGMCVSRATILGAATKYATVRRQASV  
 PGSAEVANAREDECVRLFNRPQSLTSEGGADAEALGVYVETIKTEGPDASGRNNAWY  
 PCSIETLADYVYITAEFGSPITLPBILASQCYVDAGSNCTIYAPSDNGRHLPTOTTNPRF  
 PAGGCVKRAIDLVTYVAEGRDPAAGSTIVQLLGKVAQVKEPAA"  
 complement(6763..11235)

CDS	complement(6763. .11235)
-----	--------------------------

```

/note="Identified by sequence similarity; putative; ORF
located using Blastx/Glimmer/Genemark"
/codon_start=1

```

```

/transl_table-11
/product="glutamate synthase alpha subunit"
/protein_id="AAM34925.1"

```

translation="MAPRTROGLNDGLYSVRDERDACGFGVAQLDDQPSRALVDI  
AIALSRMTHRGVAADGLTGDGCLLRK<sup>2</sup>DAFLRGLARDAGIT<sup>1</sup>LGTIRYAAGVFLP

QTEDEFTLALFLARRAEQALHAVEHFVYTLSPNGISYKGMVLPDKLSTFYPDLORSS

YEFYGLNTEPMDGPAGIVACDSRYAACMLDRNGLRPARMWLTSDRHFLVASSEAGWELL

DP SLVEEPPSEOTLR SYHKLFLQSLTEEVEQVLRPLAETEQEATGSMGDDTPMAVLRSQ  
 TRPLDYFRQAFQAVTNPPIPLREGIAMSITLTOISRETNI FHAGAE TVNHVILNSPYV

RYRVPDRPMALLATGAVINHHCKVGLRQVNI IETGTARDPHMACLLGVGATAV

YVHGGEYHMYNPDPVMTLQRASTRGTGADGDMQKYVDAVHARPASTLRDLVOLKRAADPTPT

EKRSKIKQVASGREGVTPPEYLVAEVLQIKVAQGAKPGEQQLPGHKVNELIARLRAYA  
KRGIGLISPPBHDTYSIEDIADLTVDIKONBTAIUSVKIVAHAGVSTAAQVYKAC

GLDVKKALLGADSFGEFTAPMIVLGCKYLRICHLNNCATG VATQDERLRAGYFTG LPE

QYDGSHCMAORLYESPDSLATOMDGLLADALANKSGGDHRLIHTNDRSICARLSGAI  
 ARAHGNIHGMSDAPLNLRFRTAGOSFGAFNAGGLOLEGEANDYVGKGMAGGRLVFVF

YMTDGVLLVGLKVGLEFGAGFTTGLGLAVLDIERDFVDRYNHELIDIRVSAEGFENHRH  
QUTYMTTSPUDPTWCTSTWAOOTDEEDNDVICKETATVINDAACTCTMECTDPA#

a 3052 C 3/43 g 20/1 L

9.78; Score 128; DB 1; Length 11366;  
51.48; Pred. NO. 9.7e-14;

[illegible]

CCGCTGGACGAGGCCCAATTGGGGCTGATCAGAGAATCGCTCGAGGCACC 2900

GGGGCCTACTGCATGATTGACATGCACAACCTTGGCCCGCTACACGGGG 383

[illegible]

ACCAGCAGCTGCCGGCCGGCGGTGGCCGATCTGTGGCCGCCGGCTGGC 2780

OY	444A	AAAGTACTACGAGCAACGACAAATATATCTTTGGCCGTGATGAACGAGCCGACAGCT	503
Db	2779	GCTGGATTTCAAGGACGACAAAGGCGGTGATCTTCGGCGTGAATGAACGAGCCCAATGGCAT	2720
OY	504	CGACATTTAGATCTTGGGGCGAGACGCTGGCCAAAAGGTCGTCACTGGGATCGAAAGCCG	563
Db	2719	TTTCGGCGGCCGAGCTGGGCCAATGCCGCGCAGGGAGACATCACTGGGATCCGCAAGCCG	2660
OY	564	CGCCACCTGCGAGATGATCTCTGCCCCGGGAACCAATTTGCCAGGTCGAGACGTATGT	623
Db	2659	CGCGA---AGATCTGATCTGTGGTCCGGGCGACCGCTTACACCGGCGCACACGCTGCGC	2603
OY	624	GTCACCTGGCAGCGGGGAAAGCCCTGGGCAAGTTACG---AACCGGATGGAAGACCGA	680
Db	2602	CAGCACACGACTAGCGCGTCTCCAAAGCAAGAGCGCTGGAATATCTCMAAGACCCGGTAA	2543
OY	681	TTTGCTGATCTTGGATGTGTCACACAGTATCTCACATCAACAACTCGGGTCGACGCCGA	740
Db	2542	CAACCTTGCTTGCATCGAAGGCGCATCACTACCTGGATTAAGAGCTACAGGGGACCAAGCGGT	2483
OY	741	GTCACACAGCAACAGTGCAGCGCTTCAC----GACTTCGGGAGATGCTGTGAGCA	794
Db	2482	ATGCACACAGCGATTCGTGATGGCCAAAGAAAGCTGCAGGGCTTCACTCTGTGCTGGCGCA	2423
OY	795	GAAACAGGCGCCAGGCATATCTCCGAAACGGGCGGTGCATGGAACCTTCGTGCATGAC	854
Db	2422	GAAACAGCAAGAAAGGCTTCTCTGGGGAAATTCGCCACGGCCAAACATCCGCTGTGGACAA	2363
OY	855	TGCCCTTCGCGCCAGAAACAAAGGCCATTAGCCAAAACACGACGCTACATTGGCTTTGT	914
Db	2362	GCGCGCTGGAAGCATGTGACTACATGATAAAGAACACGACGTTGTGCTGGCTGGAC	2303
OY	915	GGGCTGGGGTGGCGGACGTTTGAACAGTGCATCTTGACTCTGACCTCCCTGGGCA	974
Db	2302	CTGGTGGGCGGGGCGGCGTGTGGAAGCGGACTATCCCTTCACCGTCAACCGGGCAA	2243
OY	975	GCCCGCAACTACA 988	
Db	2242	GGACGGCAAGCGACA 2229	

## RESULT 9

### PSEEG

**DEFINITION** *Ralstonia solanacearum* beta-1,4-endoglucanase (egl) gene, complete cds

ACCESSION	M84922	M29098
VERSION	M84922.1	GI:151207

SOURCE	ORGANISM
Ralstonia solanacearum.	Ralstonia solanacearum

Ralstonia.

AUTHORS	Huang, J. Z., Suvoirdhaman, M. and Schell, M. A.
TITLE	Excretion of the egl gene product of <i>Pseudomonas solanacearum</i>

**MEDLINE** 89291722  
**PIRMEP** 2738021

**AUTHORS**  
Huang, J. Z. and Schell, M. A.

sequences in extracellular export of endoglucanase Egl from *Pseudomonas solanacearum*

MEDLINE 92138626  
PTMED 1735723

COMMENT	on rev 8, 2002 this sequence replaced g1:34204/.
FEATURES	Location/Qualifiers

```
/organism="Ralstonia solanacearum"  
/strain="Aw"
```



repeat\_region predicted by Homology"  
/evidence-not\_experimental  
202. .218  
/note="RS05849  
HYPOTHETICAL PARA BOX  
predicted by Homology"  
/evidence-not\_experimental  
complement(227. .243)  
/note="RS05850  
HYPOTHETICAL PARA BOX  
predicted by Homology"  
/evidence-not\_experimental  
complement(263. .279)  
/note="RS05851  
HYPOTHETICAL PARA BOX  
predicted by Homology"  
/evidence-not\_experimental  
complement(307. .323)  
/note="RS05852  
HYPOTHETICAL PARA BOX  
predicted by Homology"  
/evidence-not\_experimental  
complement(374. .390)  
/note="RS05853  
HYPOTHETICAL PARA BOX  
predicted by Homology"  
/evidence-not\_experimental  
complement(516. .532)  
/note="RS05854  
HYPOTHETICAL PARA BOX  
predicted by Homology"  
/evidence-not\_experimental  
complement(670. .686)  
/note="RS05855  
HYPOTHETICAL PARA BOX  
predicted by Homology"  
/evidence-not\_experimental  
complement(733. .749)  
/note="RS05856  
HYPOTHETICAL PARA BOX  
predicted by Homology"  
/evidence-not\_experimental  
complement(853. .869)  
/note="RS05857  
HYPOTHETICAL PARA BOX  
predicted by Homology"  
/evidence-not\_experimental  
complement(919. .935)  
/note="RS05858  
HYPOTHETICAL PARA BOX  
predicted by Homology"  
/evidence-not\_experimental  
complement(1219. .1244)  
/note="RSP0001; RS02238"  
1219. .1244  
/gene="repa"  
/function="elements of external origin; plasmid-related  
functions non-bacterial functions"  
/note="Product confidence : probable  
Gene name confidence : probable  
predicted by Codon\_usage  
predicted by Homology  
predicted by Framed"  
/codon\_start=1  
/evidence-not\_experimental  
/transl\_table=11  
/product="PROBABLE REPA REPLICASE PROTEIN"  
/protein\_id="CAD17152.1"  
/db\_xref="GI:17430468"  
/translation="MPGRIVPSGGVATPPEKFORLEDEAGAMEREDAMSGEYVGL  
SRASVOTLPYRAPKSPVWRRSSGNISLMIOPGYFTQORSEASNGRORIVSEVS  
FGPYGSPRLAMWIKREIMAKKKRGEQGVDEDRISLGSLSDFMNLGIPMATG

gene complement(3120. .3323)  
/gene="cspd3"  
/note="RSP0002; RS05790"  
complement(3120. .3323)  
/gene="cspd3"  
/function="miscellaneous; not classified regulator"  
/note="Product confidence : probable  
Gene name confidence : putative  
predicted by Codon\_usage  
predicted by Homology  
predicted by Framed"  
/codon\_start=1  
/evidence-not\_experimental  
/transl\_table=11  
/product="PROBABLE COLD SHOCK-LIKE TRANSCRIPTION REGULATOR  
PROTEIN"  
/protein\_id="CAD17153.1"  
/db\_xref="GI:17430469"  
/translation="MAGTVKWFNETKGGFTTPDGGADLFAMHSEIQQSGFKTKD  
GOKVSEVQKQKGLQASGIRPE"  
3728. .4396  
/gene="para"  
/note="RSP0003; RS01975"  
3728. .4396  
/gene="para"  
/function="elements of external origin; plasmid-related  
functions non-bacterial functions"  
/note="Product confidence : probable  
Gene name confidence : putative  
predicted by Codon\_usage  
predicted by Homology  
predicted by Framed"  
/codon\_start=1  
/evidence-not\_experimental  
/transl\_table=11  
/product="PROBABLE PARTITION PROTEIN"  
/protein\_id="CAD17154.1"  
/db\_xref="GI:17430470"  
/translation="MSAKIVYFNOKGCGKTTVSMHAGTLGVRSKTLVDMDKDG  
TATRNAAOPEEKPPASVYGLPAGAMHRRFIDYIVVDCPPAVSPASS  
ALLISDIATIPVPSBPDLMVAKAQAQOVNERTLRVNLNMORVSLAKOA  
IETLDDGDIPLEMISMISRSFRECQALGATVHGVPARAEINEVDIVNVEVDLDR  
EQQA"  
4393. .5397  
/gene="parb"  
/note="RSP0004; RS01976"  
4393. .5397  
/gene="parb"  
/function="macromolecule metabolism; macromolecule  
synthesis, modification; dna replication, repair,  
rest./modif."  
/note="Product confidence : putative  
Gene name confidence : putative  
predicted by Codon\_usage  
predicted by Homology  
predicted by Framed"  
/codon\_start=1  
/evidence-not\_experimental  
/transl\_table=11  
/product="PUTATIVE PARTITIONING PROTEIN"  
/protein\_id="CAD17155.1"  
/db\_xref="GI:17430471"  
/translation="MSAKLSLKDGLAGIAEKTAKAPVDREALAEQATYSHPRGL  
GQAPAPASASQAEFSRRRVKTIPLERLDENPLNARIYVDOIVQERASIAITHGO  
QTPGLAAPPAKPGWYVLIDGIRKRALAAAKAECEPIEGSLDIDFRLSFVLE

Query Match 7.7%; Score 101; DB 1; Length 216050;  
Best Local Similarity 48.0%; Pred. No. 7.2e-09;  
Matches 324; Conservative 0; Mismatches 345; Indels 6; Gaps 1;

QY 261 CACGGTGACGGCAGCTGACAGCTCACTGGGGCTCTACAACAAGTCTGTCACGC 320  
 Db 189097 CACGCTCAACAGGCGCTCTGACAGGAGAGCTGCGGCTGACCGGCTTCTGTCACGC 189156  
 QY 321 CTGCTGACAGAGGGGCGCTCTGCTGATGATTGACATGCAACAATTGGCCCTACACGC 380  
 Db 189157 CGTGAAGCGGCGCGCCGACAGGAGTGTGCTGATCCGACACAATGACGCGCTACTACGC 189216  
 QY 381 CGGATCATCTGGCCAGGAGGCGTGTGCGAGACATCTTGTGCGCTCTGCTGTCACAT 440  
 Db 189217 CACGCTGATGCGCTGACAGCGGCGGTGCTCCACACAGCGCTACCGCTTCTGCGGCGCT 189276  
 QY 441 CGCAAACTACTACGAGACAGACAGACAGATCATCTTGGCCCTGATGACAGCGCCACGA 500  
 Db 189277 GCGCACCCAGTTCAAGGACCAATGCGCGCTCATCTTGGGCTGATGACAGAGCCCAATTTC 189336  
 QY 501 CCTGACATTTAGATCTGGGCGGACAGCTGCCAAAAGTCTGCTCACTGCTGATCCGAAGGC 560  
 Db 189337 GATGCGCGACCGAGAGTGGCTGTCCGGCGCCACAGCGCGCTGCGCAATCCGCTCCGC 189396  
 QY 561 CGGCGCCACTTCGAGATGATCTCTGCTGCGGAAACCACTTTGCGCAGCTGCGAGACGTA 620  
 Db 189397 CAATGCGAGCAACGTGTGTCTGCTGCGGCAAGCTTGCAGACCGGATGGAAGCACCGA 189456  
 QY 621 TGTGTCTCACTGGCAGCGCGGAGAGCCCTGCGAAGATTGACAGCCGATGGAAGCACCGA 680  
 Db 189457 CCAGAACTGTGATGCGGACGCGGACAGGACCGCTCATGAAAGGCAATGACACCGGCGC 189516  
 QY 681 TTTGCTGATCTTATGATCTGCAACAGTATCTGACATCAACAATCTCGGGTGGCACGCCGA 740  
 Db 189517 CAACTCTCTTTCGAGGTGACAGTACAGTCTGACGCGGCTTCTGCGGCAAGTGGCGCAG 189576  
 QY 741 GTGACACAGACAGCAACGTGACGCC-----TTCAGCACTTTCGCGGACTGCTGAGGCA 794  
 Db 189577 CTGGGTGAGGCGCACACATGCGGCGCGGACGCGCTGAGAGACTTCACCACTGGCTGCGCAG 189636  
 QY 795 GAACAAGCGCGAGGCGCATCTCTCGAAGAGCGCGCTGTCATGGAACCTTCTGTCATGAC 854  
 Db 189637 CAACGCGCTATGCGGCTTCTGCGGAGTGTGCGGCGGCTTCAACAGACACTGCAACCA 189696  
 QY 855 TGCCTTTCGCGCGCAGAACAGAGGCGCATGAGGAACAGAGCGCTCATTTGGCTTTT 914  
 Db 189697 AGCGGTTCACCAATGCTCTACCTTGTGTGAGAACATGCGATGTGTGAGCGGTGGCG 189756  
 QY 915 GGGCTGGGGTGGCGG 929  
 Db 189757 GTGGTGGGCGGGCGG 189771

RESULT 11  
 AF440003 1185 bp mRNA linear PLN 30-MAY-2002  
 LOCUS Talariomyces emersonii. endoglucanase mRNA, complete cds.  
 DEFINITION AF440003  
 ACCESSION AF440003.2 GI:21264633  
 VERSION  
 KEYWORDS  
 SOURCE  
 ORGANISM  
 Talariomyces emersonii.  
 Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;  
 Eurotiales; Trichocomaceae; Talariomyces.  
 REFERENCE 1 (bases 1 to 1185)  
 MURRAY,P.G., COLLINS,C.M. and TUOHY,M.G.  
 Molecular cloning and expression of endoglucanase genes of  
 Talariomyces emersonii  
 JOURNAL Unpublished  
 REFERENCE 2 (bases 1 to 1185)  
 MURRAY,P.G., COLLINS,C.M. and TUOHY,M.G.  
 Direct Submission  
 JOURNAL Submitted (26-OCT-2001) Biochemistry, National University of  
 Ireland, Galway, Newcastle Road, Galway, Ireland.  
 REFERENCE 3 (bases 1 to 1185)  
 MURRAY,P.G., COLLINS,C.M. and TUOHY,M.G.

TITLE Direct Submission  
 JOURNAL Submitted (30-MAY-2002) Biochemistry, National University of  
 Ireland, Galway, Newcastle Road, Galway, Ireland  
 REMARK Sequence update by submitter  
 COMMENT On May 30, 2002 this sequence version replaced gi:17017347.  
 FEATURES  
 source  
 1. 1185  
 /organism="Talaromyces emersonii"  
 /db\_xref="taxon:68825"  
 37. 1041  
 /EC\_number="3.2.1.4"  
 /note="glycosyl hydrolase"  
 /codon\_start=1  
 /product="endoglucanase"  
 /protein\_id="AAL3630.2"  
 /db\_xref="GI:21264634"  
 /translation="MKSTVCGLLAAGALAPAKERSIKRASPQWFGNSGAE  
 FQGNRIPEVECTDYTFPNTAIGVLDQGNITRVPFLMRATPNQRTGVSATLOG  
 YSOVINYITSHGASAVDAIDPHNRYRNNIISPSDEQTFWNTLASNFADNDVATFDTN  
 NEYHNDDESILVQNLQALDIFRAAGTSOYIFEGNSMTGATWQVNDAMNLTDP  
 QNKIIVEMHOYLDSDSGSTSDQCVNATIGQDRAVSATAMLKQNGKRAIIEFPAGANS  
 VCSAVTGLDLADNTDVTGAIIMWAGPMWASYSIFSPSPGIAVEQLPLQPLVLT  
 "

BASE COUNT 275 a 324 c 344 g 242 t  
 ORIGIN  
 Query Match 7.0%; Score 92.2; DB 8; Length 1185;  
 Best Local Similarity 48.5%; Pred. No. 4.1e-07;  
 Matches 373; Conservative 0; Mismatches 363; Indels 33; Gaps 3;

QY 205 GACGAGCGCTCAACGCTCTTGCATATCGCTACATGCAAGTTGTCTCAACACAG 264  
 Db 247 GACCAAGGCGATGAAATATCTTCGCGTCCGTCTGATGAGGCCATGTCGCCAACAG 306  
 QY 265 GTGACGGCGAAGCTGGAGCGAGCTCAACTAAGTGGGCTCTCAACAAGGTGTCACGCCGT 324  
 Db 307 ATGACGGCGCCCGTGTGATTCGCGTACCTGCAAGGCTACACAGGTGATCACTATT 366  
 QY 325 CTCGAGACGGCGGCTTACTGATGATGACATGACACACTTTGCCCGCTACACAGCGCGC 384  
 Db 367 ACCAGTACGGGGGCGGTGCGGCGGTGATGACCCGATACATGACGATACACAAAT 426  
 QY 385 ATCATCGCGCCAGGAGGCGGTGCGAGACATCTTTGTGCACCTCTGGGTCAATCCCA 444  
 Db 427 ATCATCTTCC-----TCGCCGTGCGACTTCTGGAACATATTCCG 474  
 QY 445 AAGTACTGAGAGCAACAGCAAGATCATCTTTGGCGTATGAGAGAGCCGACACTC 504  
 Db 475 TTCACCTTTGGCGCAATGACGGCGTATTTCACACGAAACAGATACACGACATG 534  
 QY 505 GACATTGAGATCTGGGCGCAGACGTGCCAAAAGTCTGCTGATGCGATCCGAAGCGCGC 564  
 Db 535 GACGAAGCGCTCGTGTGACACTCAACAGCGCGCATGACAGGCGATCCGGCGGCGC 594  
 QY 565 GCCACCTGCGAGATGATCTCTGCGCGGAACCAACTTTGCCACGCTGAGAGCTATGTG 624  
 Db 595 GCCACGTGCGAGTACTATTCTGTGAAGGAACCTGTGAGCGGCGCTGACCT----- 649  
 QY 625 TCCACTGGCAGCGGGAAGCGCTGGCAAGATTGCAACCGCGATGGAAGCAGCATTTG 684  
 Db 650 -----GGAGCGAGGTAAAGCAGCGCGTGGCAACCTGACAGATCTCCAGAACAG 699  
 QY 685 CTGACTTTGATGTGTCACAAAGTATCTGACATCAACAACCTCGGGTGGCAGCGCGAGTGC 744  
 Db 700 ATCGTTTACGAGATGACACATGACTGCACTGCGAGCGGTGGGACATCGACACAGTGC 759  
 QY 745 ACCACAGCAACGTC-----GACGCTTCAACGACTTCCGGAGCTGAGCAGAC 798  
 Db 760 GTCAACGCGACATCGGGCAGAGACCGGCTGCGCCACAGCGCGCTGGAACAGAAC 819  
 QY 799 AAGCGCCAGCGCATCATCTCGGAAGCGCGCGCTGCATGGAACCTTGTGATGATGCTGC 858  
 Db 820 GGCAGAAAGCGATTTGGCGGAGTTCGCGGCGGTGCGAAGCAGCTGTGCGAGTCCGCC 879

OY 859 TTCTGGCCCGAGAACAGCCATTAGCAAAACAGCAGCTTACATGGCTTTGGGC 918  
 11  
 Db 880 GTACAGGGGCTGCTCAGCATCTGGCCGATTAACGATGTCTGGACGGGGCTATCTGG 939  
 OY 919 TGGGGTCCCGCAGCTTTGACAGCTGATCATCTGACTGTGATCCCC 967  
 11  
 Db 940 TGGGCTGCTGGCCGCTGTGGGCAAGCTACATCTCTCCATGGAGCCGC 988  
 RESULT 12  
 AF233448 2010 bp DNA linear BCT 12-JUN-2002  
 LOCUS Sinorhizobium meliloti endoglucanase precursor (ends) gene,  
 DEFINITION complete cds.  
 ACCESSION AF233448  
 VERSION AF233448.1 GI:12005273  
 KEYWORDS  
 ORGANISM Sinorhizobium meliloti.  
 Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;  
 Rhizobiaceae; Sinorhizobium.  
 REFERENCE 1 (bases 1 to 2010)  
 AUTHORS Michaud, P., Belatch, A., Courtois, B. and Courtois, J.  
 TITLE Cloning, sequencing and overexpression of a Sinorhizobium meliloti  
 JOURNAL M5N1CS carboxymethyl-cellulase gene  
 MEDLINE Appl. Microbiol. Biotechnol. 58 (6), 767-771 (2002)  
 PUBMED 12021797  
 REFERENCE 2 (bases 1 to 2010)  
 AUTHORS Michaud, P., Belatch, A., Courtois, B. and Courtois, J.  
 TITLE Direct Submission  
 JOURNAL Submitted (10-FEB-2000) IUT d'Amiens (GB), LBM, Avenue des Facultes  
 / Le Bailly, Amiens 80025, France  
 FEATURES  
 source  
 1. 2010  
 /organism="Sinorhizobium meliloti"  
 /db\_xref="taxon:382".  
 gene 505..1599  
 /gene="ends"  
 CDS 505..1599  
 /gene="ends"  
 /note="carboxymethyl cellulase"  
 /codon\_start=1  
 /transl\_table=11  
 /product="endoglucanase precursor"  
 /protein\_id="AAG4364.1"  
 /db\_xref="GI:12005274"  
 /translation="MKSMTPTARKGMLPLVRAGLALTLTATTSOAVLASGCLRG  
 INLAGEFEEDVEFGATVYPSDEIRFAKGFNSVRLPFSMSRLTSLNADPAA  
 EPERLKDTYRLREAGQIVYLDPHNAYARKGLISEAVPYRAFDFNSKLSLARGNO  
 DGIARGLAEPHTMPTBQMLTANVAIAINSTGAKNLLIVGNSVSGAHSMWGEDYG  
 GANGVVNLGKDPDLHYAFVHQLDVEFSGRDKSRAEDAIAALENTQWLRLDNGK  
 RGYLQFQVPRGDEACVOALSRLVDEVERDWMISWAYAAGDMWPAEELNIQPTDA  
 GDRPOLGLSRLSDPSFQAMTCSLSGG"  
 sig\_peptide 505..585  
 /gene="ends"  
 mat\_peptide 586..1596  
 /gene="ends"  
 /product="endoglucanase"  
 BASE COUNT 368 a 599 c 625 g 418 t  
 ORIGIN  
 Query Match 6.8%; Score 89.8; DB 1; Length 2010;  
 Best Local Similarity 48.5%; Pred. No. 1.1e-06;  
 Matches 374; Conservative 0; Mismatches 382; Indels 15; Gaps 4;  
 OY 173 ATGGCCCGCGCCGATGATGATTCGCGAAGACGCGCTCAAGCTTTGGCATAT 232  
 11  
 Db 698 ATCCCTCCAGACGACCATCTGCTATTTGCGACAAAGGCTTCACTCGCTTGC 757  
 11  
 OY 233 CCCTCATGCGAGTTTGTCTCTCAACACAGCGTGGACGCGCAAGCTGACGACTCAACT 292  
 11

Db 758 CCTTTTCCTG---GTGCGGCTCGACAGCTCTCTCAAGCGGCACTTGATGGCGGTGAAT 814  
 OY 293 GGGGCTCTACACAAAGTGTCAAGCCCTGCTCAGANCGGGCGGCTATGATGATG 352  
 11  
 Db 815 TCGAGCGGCTGAAGATACGCTGCTGAGCTGCGGAGGCGGGGAGAGANTGTTGTTCTG 874  
 OY 353 ACATGCAACACTTTGGCCGCTTACACAGCGGCGATCACTGCGGAGGAGGCTGCGAG 412  
 11  
 Db 875 ATCCCTACAACTATGCCCGTACCGGCGAGTGTATGCGCTGGGAAGCCGCTTACG 934  
 OY 413 ACATCTTTGTGACCTCTGGGCTCCAGATCGCAAGTACTACAGAGAACGACAGATCA 472  
 11  
 Db 935 AACCTTTGGCGGATTTCTGAGCAAGCTCAGCTCCGCTTGCGCAATACAGAGCGCATCG 994  
 OY 473 TCTTGGCCCTGATGACAGACCGCCAGACCTCGACATGTGATGATGCGCCAGACGTCG 532  
 11  
 Db 995 CGTTGGATGATGACAGACCGCCGATACCAATCCAGCGGAGAGAGTGT---GCTACCGGGG 1051  
 OY 533 AAAAGTCTGCTACGCGATCGGATCGGAAAGCGCGCCACCTGCGAGATGATCTCTGCGG 592  
 11  
 Db 1052 CGAATGTCGGATCGCGCCGATCGGATCGGATCGGATCGGATCGGATCGGATCGG 1111  
 OY 593 GAACCAACTTTGCGACGCTGAGACGCTATGTGTCACTGCGACCGCGGAAGCCCTCGCA 652  
 11  
 Db 1112 GAAACTCTGCTGCGGAGACATAGCTGATGGCGAGGAGCTATGTGGCGGCAACGCGG 1171  
 OY 653 AGATTACGAC---CCGGATGGAAGACCGATTTGCTGCTACTTTGATGTCCAAAGATTC 709  
 11  
 Db 1172 TTGTCTAGCTGCTGCTGGAAGATCCGCTGATATATATGCTGAGCTCACAGATATC 1231  
 OY 710 TCGACATCAACACTCGCGGCTGCGACCGCGAGTGA-----CCACGACACAGTGGAG 763  
 11  
 Db 1232 TGAACGATGCTCTTTTCCGGACCAAGACATTTGACCGCGGAGAGATGCTATGCGG 1291  
 OY 764 CCTTCAACGACTTCGCGAGCTGCTGAGCGACAAACAGCGCCAGCCATCATCTCGAAA 823  
 11  
 Db 1292 CAATCGAAGACTACACGACGACTGCTCGACACATGGAAGCGCGGCTACCTGGCCAGT 1351  
 OY 824 CGGCGCGCTCATGGAACCTTCTGCTGATGCTGCTCTTCTGCGCCAGAACAGCCATTA 883  
 11  
 Db 1352 TCGCGCTGCGAAGATGAGAGCCCTGCTGAGCTCTCACTGCTGATGATGATGATG 1411  
 OY 884 GCGAAAACGACGCTTACATGCTTTGGCGCTGGGCTGGCGGAGCT 934  
 11  
 Db 1412 AACGGGGCGGAGTGTCTGATGATGATGATGATGATGATGATGATGATGATGAT 1462  
 RESULT 13  
 AE004077 9997 bp DNA linear BCT 15-JUN-2001  
 LOCUS Xylella fastidiosa 945c, section 223 of 229 of the complete genome.  
 DEFINITION AE004077 AE003849  
 ACCESSION AE004077.1 GI:9107952  
 VERSION  
 KEYWORDS  
 ORGANISM Xylella fastidiosa 945c.  
 Bacteria; Proteobacteria; gamma subdivision; Xanthomonas group;  
 Xylella.  
 1. (bases 1 to 9997)  
 REFERENCE Simpson, A.J., Relbach, F.C., Arruda, P., Abreu, F.A., Acencio, M.,  
 Alvarenga, R., Alves, L.M., Araya, J.E., Bala, G.S., Baptista, C.S.,  
 Barros, M.H., Bonaccorsi, E.D., Bordin, S., Boye, J.M., Briones, M.R.,  
 Bueno, M.R., Camargo, A.A., Camargo, L.E., Carraro, D.M., Carrier, H.,  
 Colauto, N.B., Colombo, C., Costa, F.F., Costa, M.C., Costa-Neto, C.M.,  
 Coutinho, L.L., Cristofani, M., Dias-Neto, E., Docena, C., El-Dorri, H.,  
 Facincani, A.P., Ferreira, A.J., Ferreira, V.C., Ferro, J.A.,  
 Fraga, J.S., France, S.C., Franco, M.C., Frohme, M., Furlan, L.R.,  
 Gantier, M., Goldman, G.H., Goldman, M.H., Gomes, S.L., Gruber, A.,  
 Ho, P.L., Hohnsels, J.D., Junqueira, M.L., Kemper, E.L., Kitajima, J.P.,  
 and Marino, C.L.  
 The genome sequence of the plant pathogen Xylella fastidiosa. The  
 Xylella fastidiosa Consortium of the Organization for Nucleotide  
 Sequencing and Analysis

JOURNAL Nature 406 (6792), 151-157 (2000)  
 MEDLINE 2036517  
 PUBMED 10910347  
 REFERENCE 2 (bases 1 to 9997)  
 AUTHORS Simpson A.J.G., Relnach F.C., Arruda P., Abreu F.A., Acencio M., Alvaranga R., Alves L.M.C., Araya J.E., Bata G.S., Baptista C.S., Barros M.R., Bonaccorsi E.D., Bordin S., Bove J.M., Brites M.R.S., Bueno M.R.P., Camargo A.A., Camargo L.E.A., Carraro D.M., Carter H., Colauto N.B., Colombo C., Costa F.F., Costa M.C., Costa-Neto C.M., Coutinho L.L., Cristofani M., Dias-Neto E., Docena C., El-Dorri H., Facincani A.P., Ferreira A.J.S., Ferreira V.C.A., Ferro J.A., Fraga J.S., Franca S.C., Franco M.C., Frohme M., Furian L.R., Garnier M., Goldman G.H., Goldman H.S., Gomes S.L., Gruber A., Ho P.L., Hohlseil J.D., Junqueira M.L., Kemper E.L., Kitajima J.P., Krieger J.E., Kuramae E.E., Laigret F., Lambas M.R., Leite L.C.C., Lemos E.G.M., Lemos M.V.F., Lopes S.A., Lopes C.R., Machado J.A., Machado M.A., Madeira A.M.B.N., Madeira H.M.F., Martino C.L., Marques M.V., Martins E.A.L., Martins E.M.F., Matsukuma A.Y., Menck C.F.M., Miracca E.C., Miyaki C.Y., Monteiro-Vitorello C.B., Moon D.H., Nagai M.A., Nascimento A.L.T.O., Netto L.E.S., Nham J.A., Nobrega F.G., Nunes L.R., Oliveira M.A., de Oliveira M.C., de Oliveira R.C., Palmeri D.A., Paris A., Pelxoto B.R., Pereira G.A.G., Pereira J.H.A., Pesquero J.B., Quaggio R.B., Roberto P.G., Rodrigues V., de M. Rosa A.J., de Rosa J.V.E., de Sa R.G., Santelli R.V., Silva Sawaaki H.E., da Silva A.C.R., da Silva F.R., da Silva A.M., Silva Jr. W.A., da Silveira J.F., Silvestri M.L.Z., Siqueira M.J., de Souza A.A., de Souza A.P., Terenzi M.F., Truffi D., Tsai S.M., Tsubako M.H., Vallada H., Van Sluys M.A., Verjovski-Almeida S., Vettore A.L., Zago M.A., Zatz M., Zeldin J. and Zebub J.C.

TITLE Submitted (02-JUN-2000) Organization for Nucleotide Sequencing and Analysis, Bioinformatics Lab - IC/Unicamp, C.P. 6176, Campinas, SP 13083-970, Brazil

FEATURES location/Qualifiers

1..9997

/organism="Xylella fastidiosa 9a5c"

/db\_xref="taxon:160492"

/clone="9a5c"

144..1214

/gene="Xf2708"

144..1214

/gene="Xf2708"

/note="similar to SP1P17974 (percent identity: 47 %/query alignment coverage: 88.8 %/subject alignment coverage: 74.2 %); identified by sequence similarity; putative; ORF located using Glimmer/RBSfinder"

/codon\_start=1

/transl\_table=11

/product="endo-1,4-beta-glucanase"

/protein\_id="AAE85505.1"

/db\_xref="GI:9107953"

/translation="MPSFLCPICPSCFLILLTALTTPASPASHATNLTLYAGINL GPSPSPSKAGVLYKDYLYSASDVYVKKGNVRLPLMERVQPTLNDLADQ LKLIIRAVQAKKQIITIDVYNNYSYNDLICTDVPISAFADLRKRSLOSPANK AVTFGLMNEHETSTTWAQQAQALNAIRSTGCNNLVPGTASGASWSLSTYK IPRGEALNTNDPAKHAFAEYHQVLYNENSTGTGECISTIGAELEAFNMTATYRK TGLGEFATGNTNDQALQEGMLSYIEENADVWLGLTWMSNPFQDPYFNLHPNSD GSDPKQMSILESARAKTKIKH"

complement(2196..3680)

/gene="Xf2709"

complement(2196..3680)

/gene="Xf2709"

complement(2196..3680)

/gene="Xf2709"

/note="similar to SP1P09832 (percent identity: 55 %/query alignment coverage: 97.4 %/subject alignment coverage: 101.9 %); identified by sequence similarity; putative; ORF located using Glimmer/RBSfinder"

/codon\_start=1

/transl\_table=11

/product="glutamate synthase, beta subunit"

/protein\_id="AAE85506.1"

/db\_xref="GI:9107954"

/translation="MSRKAHFOFLDLPROMPQRIYELKRTSGMRELKYGDKAEAOY

gene

CDS

OAGRLDGCNPGYCSWKCVPVHNAIPQWLQVQENRIDEAALCHANTNPEVCGRCVQ DRLCESCTLEEGFAGVYIGAVERYIVDTALQDGRPDLSHVQPTGMVAAGVAGPL ACADRILAGVAVYVDRYDQIGLOFGFSPLEADKAVSTRKYLEGVEFELCV EIKDGIETALLEHYADVFLGVGYRTDGLGGODLKNVLPALPFLYONGRIVSND PGRPIAGMEDVQLODPVTKSVYVGGGPGTDCVSAIRLCTATYCTIYRDDNAM PGSAREYANAREEVRLENRPOLSTIOSGTDKQAGVETVETLAEADKRNAPFI OGTEVLEADAVVIAEFGSPMLEPWLQAOISNSRIVALSVDSDGTLTYOTTHK LFGAGDGVADLVAVAGRGRAEISVIMLATQVMDVAVI"

complement(3821..8290)

/gene="Xf2710"

/note="similar to SP1P09831 (percent identity: 53 %/query alignment coverage: 99.7 %/subject alignment coverage: 97.9 %); identified by sequence similarity; putative; ORF located using Glimmer/RBSfinder/start codon shift: -117"

/codon\_start=1

/transl\_table=11

/product="glutamate synthase, alpha subunit"

/protein\_id="AAE85507.1"

/db\_xref="GI:9107955"

/translation="MVSSTRORIEESMLYASDERDAGFGVNAQLODDPSALLEIA IALSRTHTGGVAAADLTGDDGGLIRKPDATFLAAGAGTILGPRVAVVPLPG ERAAAROVLAELDRQVRCGREGVINSYCOLARPLPHEIYEVNDPDKH SEDRGVATLARKARQUCDHDFYATLSYATSKVMYLPDKIPVTPLOKSE LSSSVYHOFESTNLPNPLAHPLELHANEINTEGNRMAOARSVWOSPRED MAFSDPVYSKGSOSLDNMLELLANGSDLOALVILPPTRTLEFDVADLASF EFGALNTEPMKDGAGIYADGRYACMLNDGRPARMLTADHFLFVASEAVELP VERNIRKRGKPGEMMAIDKRGDLDSDIDRINRARAQKMOLOQGVYLTDEL D PSIVEEPSPDOTCSYOKLPOLSTEVEYVQMLTAECELTARANDRIAEVGLKY RPLXDFRQAFQVYTPNPPIDPLAESIMSTTODGRTNIFNAGPEVNVITDPSYL SORLROLMPOLYLEKNRLIDISTYEBELQGLKMCDEQADQANBDVYVILDSOR YPVERPVHALLATGAVVHHILRVGLRCOVNLIIEGTARDPHNAACLGATATAY PLYSOTFLNMGRRGILQKAGEOSQISYKRYGILSKITISKGICTIYASRA QLEFVIGLDDPVVALCPDPTPARIGVSLQMLTAECELTARANDRIAEVGLKY VHGEVYHNVMDVVMITLOHATRTGDAADRWMDVATRPSALROLLOLKEIPEP LDHVAEPATDVLCPDPTAISGLASPBHNAALVAMNRLOGRNSGSGGDPARTGE RSKIKOVASGRFVPEYILINAEVLDIKIQAQAKPEGGQLGPKHKNVNDILARLAK DYGILSPPHNDIYSLIEDLQIYDQVNPQALVSVKLVSHVGVETIAAGVAKA DLITVSGHDGCTGASPISSIRYAGVPELVIAEHQALVNDLREMTLODGLGK LGIDVYKALLGANSFGCTAPMIVLYGCKVLYLCHNLCATVANDREMLRNNHTGDE RVNPFRLIAEVRQWLSYLGAMSLDIIOTRTDLOOLEVSRGVYVNLSTRLLTGA YEGTHCAKRLYESPDSLOISLADANANTGCDGRFLINPRTSVTRTSGAIA RVGNHGMEDAPHLNRRTAGSGFGFVNGVGLHLEEGANDVGGKMGGLCVAR PGRAREFARTPTLGNCTCYAGGELFAAGRGERRAVANSALAVTEAGDHCCE YMTDQVLYLGVKVLNFGAGTGLAYLIDIERDVPVYNNELDIHRYVVEGEVNRQ HRLHLYVRHNLGSIWAQOILDEFROYICKFVFKPKASIESLAESTRSA"

8277..8576

/gene="Xf2711"

8277..8576

/gene="Xf2711"

/note="hypothetical protein; identified using Glimmer/RBSfinder"

similarity; putative; ORF located using Glimmer/RBSfinder"

/codon\_start=1

/transl\_table=11

/product="hypothetical protein"

/protein\_id="AAE85508.1"

/db\_xref="GI:9107956"

/translation="MLETMAAFYDTGKAATHTYSLPWNALNKPPIRLNHSSTAT HRYTCNTLRMGSSNMTDTHVHLFEYTEIHNAHNAALPHTHTLAPTRHNSAQA"

8630..9607

/gene="Xf2712"

8630..9607

/gene="Xf2712"

/note="similar to SP1P22008 (percent identity: 49 %/query alignment coverage: 86.9 %/subject alignment coverage: 97.1 %); identified by sequence similarity; putative; ORF located using Glimmer/RBSfinder/start codon shift: -123"

/codon\_start=1

/transl\_table=11

/product="pyroline-5-carboxylate reductase"

/protein\_id="AAE85509.1"

/db\_xref="GI:9107957"



```

QY 861 CTGGGCCAGACAGCCATTAGCGAAACAGCAGCTTACATTGCTTTGGGCTG 920
Db 863 GCGGGGATGCTGGATGATATGCGCCACACAGCGGATGTGGAGAGCGCGGTTGGTG 922
QY 921 GGGTGGCGGAGCTTTGACAGTGTGTACATCTGTGACTGCTCCCGGCAAGCCCG 979
Db 923 GACGGCTGGCCCCCTGTGTGGCGGATTACATGTTTCAGCATGGAGCCGCCAGCGGCGG 981

RESULT 15
AX254752 1008 bp DNA linear PAT 10-OCT-2001
DEFINITION Sequence 1 from Patent W00170998.
ACCESSION AX254752
VERSION AX254752.1 GI:16074416
KEYWORDS
SOURCE Talaromyces emersonii.
ORGANISM Talaromyces emersonii.
Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
Eurotiiales; Trichocomaceae; Talaromyces.
REFERENCE 1 (bases 1 to 1008)
AUTHORS van den Hombergh, J. P., van der Laan, J. M., Daran, J. M.,
Herweijer, M. A. and Teunel, D. P.
Talaromyces emersonii beta-(D)-glucanases
JOURNAL Patent: WO 0170998-A 1 27-SEP-2001;
DSM N.V. (NL)
FEATURES
location/Qualifiers
source 1..1008
/organism="Talaromyces emersonii"
/db_xref="taxon:68825"
1..1008
/note="unnamed protein product".
/codon_start=1
/protein_id="CAC94520.1"
/db_xref="GI:16074417"
/translation="MKRSRYVCGITAGGALAAVYKKEGKKRASPRQWNSNSGAE
FGNNITGVGEQTDYTFPNTSAIQILIDQNMIRVPLMEKVPNOMTGPVDSAYFEG
YISOVINTSHGASAVIDPHNFGRYNNILSSPSDFQTFWHTIASNFDNDNVIYFDTN
NEYHMDSESLVQNLQALIDGIRAGATISQIIEVGNSTGATWTQVNDAMAILTDP
QNKIYEMHQYLDSDSGTSDQCNSTIGODRVESATFAMLKONKRAILGEVAGANS
VCEFAVTGMGLDIYANNITDVTGAIWMAAGPMWMDYIFSMPEPSIAIEQVLPILKPYL
E"
BASE COUNT 220 a 298 c 304 g 186 t
ORIGIN
Query Match 6.0%; Score 79; DB 6; Length 1008;
Best Local Similarity 47.8%; Pred. No. 0.00011;
Matches 355; Conservative 0; Mismatches 355; Indels 33; Gaps 3;

```

```

QY 553 CGAAGGCCGGCGCCACCTCGCAGATGATCCTCTGCCCCGAGACCACTTGGACGCTC 612
Db 547 GCGCGCGGGGGCGGCACATACAGTACATCTTCGTCGAGGGCACTCGTGACGGGGCC 606
QY 613 GAGAGCTATGTGTCCATCTGCGAGCGCGGAGACCCCTCGGCAGAGATTACAGACCCGGATGGA 672
Db 607 TGGACAT-----GGACGCGAGGTCAACGACGGGATGGCGAACCTTGACGGAC 651
QY 673 AGCAGCATTTGCTACTTGTATGATGTCACAGATATCTGCATCAACATCCGGGCTG 732
Db 652 CCGCAGAACAGATGCTGTATGAGATGACCAAGTACTGGAGCGGTGGGCGACG 711
QY 733 CACGCCGAGTGCACCCACAGACAGCTC-----GACGCTTCAACGACTTGGCGGACTGG 786
Db 712 TCGGACACAGTGCATCACTGACCATCGGGCAGACCGCGTCGAGTGGCGGCGCTGG 771
QY 787 CTGAGGCGAAGAACAGCGCCAGCCATCTATCTCCGAAAGCGGCGGCTCATGGAACCTTGC 846
Db 772 TGAAGCGAAGAACGGCGAAGAGCGGCTGCGGAGTACGCTGGCGGCGCAACAGCGCTG 831
QY 847 TGCATGACTGCTTCTGGCGCCAGAACAGGCCATTAGCGAAACAGGAGCTGACATT 906
Db 832 TCGGAGAGCGGCGCTACCGGCGATGCTGACTATCTCGCCAACTACTGATGTCTGGAC 891
QY 907 GGGTTTGTGGGCTGGGCTGGCGG 929
Db 892 GTGCTATCTGTGGGCGGCTGG 914

```

Search completed: May 15, 2003, 18:46:36  
Job time : 3734.49 secs

GenCore version 5.1.4.p5.4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 15, 2003, 14:03:24 ; Search time 273.624 Seconds

(without alignments)  
10839.251 Million cell updates/sec

Title: US-10-028-245-4

Perfect score: 1317

Sequence: 1 atgcgcgaactccctctc.....acatgctgtagcgttttga 1317

Scoring table: IDENTITY\_NUC

Gapop 10.0 ; Gapext 1.0

Searched: 2185239 seqs, 112599159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

N.Geneseq\_101002:\*

- 1: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT:\*
- 2: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT:\*
- 3: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT:\*
- 4: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT:\*
- 5: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT:\*
- 6: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT:\*
- 7: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT:\*
- 8: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT:\*
- 9: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT:\*
- 10: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT:\*
- 11: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT:\*
- 12: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT:\*
- 13: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT:\*
- 14: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT:\*
- 15: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT:\*
- 16: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT:\*
- 17: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT:\*
- 18: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:\*
- 19: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT:\*
- 20: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:\*
- 21: /SID2/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:\*
- 22: /SID2/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:\*
- 23: /SID2/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT:\*
- 24: /SID2/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	289.8	22.0	1463	19	AAV29597
2	289.8	22.0	1720	20	AA15437
3	289.8	22.0	1720	24	ABL59234
4	289.8	22.0	1720	24	AA143252
5	289.8	22.0	1720	24	ABL60701
6	275.6	20.9	1849	16	AA097716
7	275.6	20.9	1849	16	AA091284
8	275.6	20.9	1849	17	AA132223
9	150	11.4	1372	21	AA53338

10	116.2	8.8	923	21	AA14886	Trichoderma reesei
11	101.4	7.7	584	21	AA14904	Trichoderma reesei
12	79.4	6.0	912	19	AA16436	Glucanase II gene.
13	79	6.0	1008	22	AA20928	Talaromyces emerso
14	74.2	5.6	985	24	AB044816	Oligonucleotide fo
15	74.2	5.6	985	24	AB044817	Oligonucleotide fo
16	55.6	4.2	390	13	AA021833	Randomising oligon
17	55.6	4.2	390	14	AA036859	PCR primer for 5'
18	55.6	4.2	390	22	AA76910	Sequence containin
19	55.6	4.2	390	24	AA172775	Oligo #7 for clonl
20	53.8	4.1	1241	17	AA113948	Bacillus VIP2a(a)
21	53.8	4.1	1241	17	AA113956	Maize-optimised-VI
22	53.8	4.1	1241	18	AA174007	Maize-optimised-B.
23	53.8	4.1	1241	18	AA174008	Maize-optimised-B.
24	53.8	4.1	1241	19	AA116184	Maize-optimised DN
25	53.8	4.1	1241	19	AA116186	Maize-optimised DN
26	53.8	4.1	1358	17	AA113949	Bacillus VIP2a(a)
27	53.8	4.1	1358	18	AA174009	Maize-optimised-B.
28	53.8	4.1	1358	19	AA116188	Maize-optimised DN
29	53.8	4.1	1389	17	AA113946	Maize-optimised VI
30	53.8	4.1	1389	18	AA174002	Maize-optimised-Ba
31	53.8	4.1	1389	19	AA116175	Maize-optimised DN
32	53.8	4.1	1399	17	AA113953	Maize-optimised VI
33	53.8	4.1	1399	18	AA174000	Maize-optimised-B.
34	53.8	4.1	1399	19	AA116173	Maize-optimised DN
35	53.8	4.1	4031	17	AA113947	Maize-optimised VI
36	53.8	4.1	4031	18	AA174010	Maize-optimised-B.
37	53.8	4.1	4031	19	AA116189	Maize-optimised DN
38	52.2	4.0	1198	18	AA163043	Aspergillus niger
39	52.2	4.0	1198	18	AA160427	Cellulase CDNA clo
40	50.6	3.8	5163	19	AA207000	Cryptosporidium pa
41	50.6	3.8	5163	21	AA207000	ORF encoding a por
42	50.6	3.8	5163	24	AB104778	C parvum GP900 gen
43	50.6	3.8	5318	19	AA120701	Cryptosporidium pa
44	50.6	3.8	5318	21	AA161848	DNA encoding a por
45	50.6	3.8	5318	24	AB104777	C parvum GP900 gen

## ALIGNMENTS

RESULT 1	AAV29597	standard; DNA: 1463 BP.
ID	AAV29597:	
XX	AAV29597:	
AC		
XX		
DT	02-SEP-1998 (first entry)	
XX		
DE	DNA sequence of the specification.	
XX		
KX	Regulatory sequence: cellulase cbh1 gene; mass production;	
KW	Humicola insolens; endo-glucanase NCE4; ds.	
XX		
OS	Trichoderma viride.	
XX		
FH	Key	Location/Qualifiers
FT	CDS	14..1453
FT		/*tag- a
FT	sig_peptide	14..76
FT		/note- "contains an intron"
FT	mat_peptide	77..1450
FT		/*tag- b
FT		/*tag- c
FT	Intron	343..525
FT		/*tag- d
PN	W09811239-A1.	
XX		
PD	19-MAR-1998.	
XX		
PF	16-SEP-1997;	97MO-JP03268.
XX		

PR 13-SEP-1996; 96JP-0243695.  
 XX (MEIJ ) MEIJI SEIKA KAISHA LTD.  
 XX Aoyagi K, Moriya T, Murakami T, Sumida N, Watanabe M;  
 XX WPI: 1998-250959/22.  
 DR P-PSDB: AAW57421.  
 XX  
 PT Regulatory sequence for *Trichoderma viride* derived cellulase cbhl  
 gene - for producing Humicola insolens derived endo-glucanase  
 XX  
 PS Disclosure: Pages 52-55; 92pp; Japanese.  
 CC The present sequence appears in the specification. The specification  
 CC describes a new regulatory sequence for *Trichoderma viride* derived  
 CC cellulase cbhl gene and the establishment of a system for mass producing  
 CC cellulase in moulds such as *T. viride*. As the regulatory sequence of  
 CC cbhl genes originating in *T. viride* can highly express objective  
 CC proteins, proteins such as cellulase can be expressed. An expression  
 CC vector containing the regulatory sequence and Humicola insolens derived  
 CC endo-glucanase NCE4 DNA was produced, and used to produce endo-glucanase  
 CC at 15 grams per litre.  
 XX  
 SQ Sequence 1463 BP; 334 A; 401 C; 367 G; 361 T; 0 other:

Query Match 22.0%; Score 289.8; DB 19; Length 1463;  
 Best Local Similarity 59.8%; Pred. No. 2e-59;

Matches 505; Conservative 0; Mismatches 337; Indels 3; Gaps 1;

QY 172 GATGGCCCGCCGATGACATTTGGCCGAGACAGCCGCTTACGTTTGGCAT 231  
 DB 593 GACGGTATCGCCGATGACACTGCTCAACGATGATGGATGATTTTCCGCTA 652  
 QY 232 TCCGCTACATGCGAGTTTGTCTCAACAACGAGTGAGCGGCACTGAGAGTCAC 291  
 DB 653 CCCGTCGATGCGAGTACTGCTGTAACACATCTGGGTGAGTCTGATCCACCACT 712  
 QY 292 TGGGCTCTCAACAACGATGCTGACGCTGTCTGAGAGGGGCGCTTACTGATGAT 351  
 DB 713 ATCTGGAAGTATGATCAAGCTGCTCAGGGGCTGCTGCTGATGATGATCATC 772  
 QY 352 GACATGACAACTTGGCCGCTTACAGCGGCGCATCTGCGGCGAGGGGCTGTGGAC 411  
 DB 773 GACATGACAACTTGGCTGCTGATGAGAGGTGATCATTTGGCGGCGCTTACAAAT 832  
 QY 412 GACATCTTTGCTGACCTCTGGGTCCAGATCGCAAGTACTGAGACAGCAACAGATC 471  
 DB 833 GCCGAGTTTACCACTCTTTGGTCCAGTGGCATCGAAGTACGGCTCTCACTCGAGG 892  
 QY 472 ATCTTGGCTGATGACAGCGCGGACGACCTGACATTTGAGATCTGGGCGAGACGTG 531  
 DB 893 TGGTTGGAATTAATGAGACCCGACGACGTGAACATCACTGCGTGGCAGCGTT 952  
 QY 532 CAAGAAGGCTGATGGATCCGAAGCGCGGCGGCACTGCGAGATGATCTCTGCCC 591  
 DB 953 CAAAGAGTCTGATGCAATCCGACGCGGCTGCTACGTCGCAATTCCTTCTGCT 1012  
 QY 592 GGAACAACCTTTGCGACGCTGAGAGCTATGTCTCACTGGCAGCGGAGAGCCCTGCGC 651  
 DB 1013 GGAATATGATTAATCAATCTGGGCGAGCTTTATTTCCGATGCGACGCGCCCTGTCT 1072  
 QY 652 AAGATTACGAACCGGATGAGACACCGATTTGCTGATCTTGTATGATCCAGATATCTC 711  
 DB 1073 CAGGTATGAGAACCTGATGATGATCAACAAGATTAATCTTGCATGTCACAGATCTTA 1132  
 QY 712 GACATCAACAACCTCGGGGTGCCAGCGAGTGCACAGACAGACAGCTGCA---GCCCTTC 768  
 DB 1133 GACTCGGACAACTCGGTACTCAGCGGAGTGCAGTCAACAACTGAGGCGGCTTT 1192  
 QY 769 AAGACTTTCGCGAGCTGAGGAGCAACAAGCGGCTGATCTCGGAACGGCG 828  
 DB 1193 GCTCTCTCGCGACTTGGCTTGCAGACAGACACCGCCAGGCTATTTGACGGAACCGCGC 1252

QY 829 GCGTCCATGGAACCTTCGTGATGACTGCTTGTGCGCCGAGAACCAAGCCATTAGCCAA 888  
 DB 1253 GGTGGCAATGTTCAAGTCTCGATCCAAAGATTGTCGCAACAGATCCATACCTCAACAG 1312  
 QY 889 AACAGCGACGCTTACATTTGGCTTTGGGCTGGGGTGGCGGACGCTTTGACACGTCGAC 948  
 DB 1313 AACTCAGATGTCTATCTTGGCTATGCTGGGCTGGGGTGGCGGCTTGCATGATGACCTTAT 1372  
 QY 949 ATCTGACTGACGTCGCCCTGCGCAAGCCGCAACTACACGCAACACAGCTCATGAC 1008  
 DB 1373 ATTCTGACGGAAGAGCTTACTGAGAGCGGTAACTGCTGGAGCGACACATCCCTAGTTAGC 1432  
 QY 1009 GAGTG 1013  
 DB 1433 TCGTG 1437  
 RESULT 2  
 AAX15437  
 ID AAX15437 standard; DNA; 1720 BP.  
 AC AAX15437;  
 XX  
 DT 07-MAY-1999 (first entry)  
 DE DNA encoding a cellulase enzyme designated SCE-3.  
 XX  
 KW Cellulase enzyme; cellulose-containing fibre; bleaching; denim-dyed;  
 KW fluff elimination; weight loss treatment; deacetylated triacetate rayon;  
 OS SCE-3; ds.  
 XX  
 OS *Trichoderma viride*.  
 FH Key Location/Qualifiers  
 FT CDS 171..1610  
 FT /\*tag- a  
 FT /product= cellulase  
 FT /note= "contains 1 intron"  
 FT exon 171..499  
 FT /\*tag- b  
 FT sig\_peptide 171..233  
 FT /\*tag- c  
 FT mat\_peptide 234..1610  
 FT /\*tag- d  
 FT Intron 500..682  
 FT /\*tag- e  
 FT exon 683..1610  
 FT /\*tag- f  
 XX  
 PN WO9854332-A1.  
 PD 03-DEC-1998.  
 XX  
 PE 27-MAY-1998; 98WO-JP02326.  
 XX  
 PR 27-MAY-1997; 97JP-0137258.  
 XX  
 PA (MEIJ ) MEIJI SEIKA KAISHA LTD.  
 PI Aoyagi K, Kono J, Kono T, Murakami T, Nakamura Y;  
 PI Sato Y, Sumida N, Watanabe M;  
 DR WPI: 1999-070218/06.  
 DR P-PSDB: AAW97208.  
 XX  
 PT Cellulose preparation containing highly active cellulase SCE3  
 PT e.g. in treating cellulose-containing fibres to enable fluff  
 PT elimination, weight loss and bleaching, and in weight loss treatment  
 of deacetylated triacetate rayon  
 PS Disclosure: Page 27-30; 44pp; Japanese.  
 XX



Db 930 GACATCCACATATGCTCGATGCAAGCGTGAATCATTTGGCCAGGAGCCCTTACAAT 989  
 QY 412 GACATCTTTGTCGACCTCTGGGTCCAGATCCAAAGTACTGAGAGACAGCAAGATC 471  
 Db 990 GCCAGTTTACAGCTTTGGTGGCAGTTGGCAGTGAAGTACCGCTTCACAGTGGAGGTG 1049  
 QY 472 ATCTTTGGCCGATGACAGACCGGACGACCTCGACATTTGAGANTTGGCCGACAGTGC 531  
 Db 1050 TGGTTGGAAATGAATGAAGACCCCGACGATGAACATCAACATTCGCTGCGCTT 1109  
 QY 532 CAAAAGTCTGCTACGTCGATCCGAAAGCGCGCCGACCTGCGACGATGATCTCTCTGCG 591  
 Db 1110 CAAAGGCTGCTACGTCGATCCGAAAGCGCGCTGCTGACGATGATCTCTCTGCT 1169  
 QY 592 GGAACCAACTTTCGACGCGTGAAGAGTATGTCGACGTCGACGCGGAGACCCCTCGGC 651  
 Db 1170 GGAATGATATATCAATGCGGCGAGCTTTATTTCCGATGGAGTGCAGCGCCCTGTCT 1229  
 QY 652 AAGATTGCAACCGGATGGAAGCAGCATTTGCTGCTGATGATGTCACAGATATCTC 711  
 Db 1230 CAGGTAAAGCACTGATGATGATCAACAGATCTTAATCTTGCATGCTCAAGTACTTA 1289  
 QY 712 GACATCAACACTCTCGGCTGCGACGCGGATGCAACAGACAGCAAGCTGCA---GCGCTTC 768  
 Db 1290 GACTCGGACACTCGGCTGCTGACGCGGATGCAACAGATGCAACATGACGCGGCTTT 1349  
 QY 769 AAGCACTTCGCGGACGCTGAGGAGCAAGACGCGGATGCTGCTGCAAGACGCGC 828  
 Db 1350 GCTCTCTGCGGCTGCTGAGGAGCAAGACGCGGATGCTGCTGCAAGACGCGC 1409  
 QY 829 GCGTCATGGAACCTTCTGATGATGCTTCTGCGGCGGAGCAAGGCTTATGCGAA 888  
 Db 1410 GGTGGCAATGTCAGTCTGCTGATGATGCTTCTGCGGCGGAGCAAGGCTTATGCGAA 1469  
 QY 889 AACAGCGAGCTGCTGATGCTTCTGCGGCGGAGCAAGGCTTATGCGAA 948  
 Db 1470 AACAGCGAGCTGCTGATGCTTCTGCGGCGGAGCAAGGCTTATGCGAA 1529  
 QY 949 ATCTTGAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1008  
 Db 1530 ATCTTGAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1589  
 QY 1009 GAGTG 1013  
 Db 1590 TCGTG 1594

RESULT 4  
 AAL43252  
 ID AAL43252 standard; DNA: 1720 BP.

AC AAL43252;  
 XX 22-AUG-2002 (first entry)  
 DE Trichoderma viride endoglucanase-related coding sequence.  
 KW Zygomycetes-originated endoglucanase; cellulose binding domain;  
 KM fibre processing; waste paper de-inking; paper pulp; ds; gene.  
 OS Trichoderma viride.  
 XX WO200242474-A1.  
 PN 30-MAY-2002.  
 PD 21-NOV-2001: 2001MO-JP10188.  
 PF 21-NOV-2000: 2000JP-0354296.  
 PR (MEIJ ) MEIJI SEIKA KAISHA LTD.  
 XX Nakane A, Baba Y, Koga J, Kubota H;  
 PA  
 XX  
 PI

XX WP1: 2002-471729/50.  
 DR Cellulose-binding domain-lacking zygomycetes-originated endoglucanase.  
 XX Cellulose-binding domain-lacking zygomycetes-originated endoglucanase.  
 PT with effect of endoglucanase activity enhanced in processing fibers,  
 PT delinking waste paper and improving freeness of paper pulp -  
 XX Disclosure: Page 90: 109pp: Japanese.  
 XX The invention comprises the amino acid and coding sequences of  
 CC zygomycetes-originated endoglucanase enzymes lacking the cellulose  
 CC binding domain. The zygomycetes-originated endoglucanase enzymes of the  
 CC invention have enhanced endoglucanase activity. The zygomycetes-  
 CC originated endoglucanase enzymes of the invention are useful for  
 CC processing fibers, de-inking waste paper and improving the freeness of  
 CC paper pulp - which is particularly applicable in detergent compositions.  
 CC The present DNA sequence represents an endoglucanase-related gene  
 CC sequence of the invention.  
 XX  
 XX Sequence 1720 BP; 390 A; 470 C; 410 G; 450 T; 0 other;  
 SQ

Query Match 22.0%; Score 289.8; DB 24; Length 1720;  
 Best Local Similarity 59.8%; Pred. No. 2.1e-59;  
 Matches 505; Conservative 0; Mismatches 337; Indels 3; Gaps 1;

QY 172 GATGCGCGCGCCAGATGAAAGCAATTCGCGGAGGAGCGGCTTCAAGTCTTCGATA 231  
 Db 750 GACGATATGCGGCGGAGATGCAAGTATGATGATGATGATGATGATGATGATGATGAT 809  
 QY 232 TCCGCTACATGCGAGTTGCTCTCAACACAGCGTGAAGCGGAGGAGGAGGAGGAG 291  
 Db 810 CCGCTGCGGATGCGAGTACCTGTAACAAATCTGCTGCTGCTGCTGCTGCTGCTGCTG 869  
 QY 292 TGGGCTCTCAACAAAGTGTCAACGCTGTCTGCGAGGCGGCTTCTGATGATGAT 351  
 Db 870 ATCTGCAAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 929  
 QY 352 GACATGCAACCTTTCGCGGCTGACAGCGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAG 411  
 Db 930 GACATGCAACCTTTCGCGGCTGACAGCGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAG 989  
 QY 412 GACATCTTTGTCGACCTCTGGGTCCAGATCCGAAAGTACTGAGAGGAGCAAGCAATGTC 471  
 Db 990 GCCAGTTTACAGCTTTGGTGGCAGTTGGCAGTGAAGTACGCTGCTCAAGTGGAGGTG 1049  
 QY 472 ATCTTTGGCCGATGACAGACCGGACGACCTGCAATTTGAGTGGGCGGACAGCTGC 531  
 Db 1050 TGGTTGGAAATGAATGAAGACCCCGACGATGCAACATGATGATGATGATGATGATGAT 1109  
 QY 532 CAAAAGTCTGCTACGTCGATCCGAAAGCGCGGAGGAGGAGGAGGAGGAGGAGGAGGAG 591  
 Db 1110 CAAAGGCTGCTACGTCGATCCGAAAGCGCGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1169  
 QY 592 GGAACCAACTTTCGACGCGTGAAGAGTATGTCGACGTCGACGCGGAGGAGGAGGAGGAG 651  
 Db 1170 GGAATGATATATCAATGCGGCGAGCTTTATTTCCGATGGAGTGAAGCGGCTGCT 1229  
 QY 652 AAGATTGCAACCGGATGGAAGCAGCATTTGCTGCTGATGATGATGATGATGATGATGAT 711  
 Db 1230 CAGGTAAAGCACTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1289  
 QY 712 GACATCAACACTCTCGGCTGAGGAGCAAGACGCGGAGGAGGAGGAGGAGGAGGAGGAG 768  
 Db 1290 GACTCGGACACTCGGCTGCTGACGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1349  
 QY 769 AAGCACTTCGCGGACGCTGAGGAGCAAGACGCGGAGGAGGAGGAGGAGGAGGAGGAG 828  
 Db 1350 GCTCTCTGCGGCTGCTGAGGAGCAAGACGCGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1409  
 QY 829 GCGTCATGGAACCTTCTGATGATGCTTCTGCGGCGGAGCAAGGCTTATGCGAA 888  
 Db 1410 GGTGGCAATGTCAGTCTGCTGATGATGATGATGATGATGATGATGATGATGATGAT 1469

[illegible]

FT mat\_peptide 535..590  
 FT /\*tag- b  
 FT /note= "endoglucanase-II coding region of exon 1"  
 FT mat\_peptide 766..1689  
 FT /\*tag- c  
 FT /note= "endoglucanase-II coding region of exon 2"  
 FT misc\_feature 433..534  
 FT /\*tag- d  
 FT /note= "linker"  
 FT misc\_feature 325..432  
 FT /\*tag- e  
 FT /note= "cellulose binding domain"  
 PN W09516360-A1.  
 XX 22-JUN-1995.  
 PD 19-DEC-1994: 94WO-EP04212.  
 PE 17-DEC-1993: 93US-0169948.  
 XX (FINN-) FINNFEEDS INT LTD.  
 PA (GEMV) GENECOR INT INC.  
 XX Bedford MR, Clarkson KA, Collier KD, Fowler T, Iaremas E;  
 PI Morgan AJ, Ward M;  
 XX WPI: 1995-231296/30.  
 DR P-PSDB: AAR79540.  
 PT New animal feed additive - comprising one or more endo:glucanase  
 PT enzymes and opt. a cello:blo:hydrolase enzyme.  
 PS Disclosure: Fig 2A-C; 92pp: English.  
 CC Endoglucanase-II (EG-II) is used in an animal feed-additive to  
 CC produce a cereal-based animal feed, especially for fowl. EG-II  
 CC may be produced recombinantly, optionally without the  
 CC cellulose-binding domain. The feed-additive can be incorporated  
 CC into a cereal-based feed (barley, wheat, triticale, rye and maize)  
 CC where it improves the conversion ratio and/or increases the  
 CC digestibility of the feed. The feed-additive enables a  
 CC conventional cereal-based feed to be modified by reducing its  
 CC energy, protein and/or amino acid content while simultaneously  
 CC maintaining the same nutritional levels of energy, protein and  
 CC amino acids available to the animal. The feed-additive also  
 CC contains cellobiohydrolase, and optionally a xylanase, protease,  
 CC mannanase or a phytase.  
 CC  
 XX Sequence 1849 BP; 444 A; 507 C; 440 G; 458 T; 0 other:  
 SQ  
 Query Match 20.9%; Score 275.6; DB 16; Length 1849;  
 Best Local Similarity 59.2%; Pred. No. 5.1e-56;  
 Matches 489; Conservative 0; Mismatches 334; Indels 3; Gaps 1;

Db 1072 GCTCAATTACAGAGCCCTTTGGTCGAGTTGGCATCAAGTACGATCTCAGTCAGGGTG 1131  
 Oy 472 ATCTTGGCCCTGATGATGAGAGCCGACGACCTCGACATTTGATGATGGGCGAGACGTGC 531  
 Db 1132 TGGTTCGGCATCATGATGATGAGAGCCGACGACCTGATGATGATGATGATGATGATG 1191  
 Oy 532 CAAGAGCTGCTACATGCGATCGGAAAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 591  
 Db 1192 CAAGAGCTGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1251  
 Oy 592 GGAACCACTTGGCAGGCTGAGAGCGATGATGATGATGATGATGATGATGATGATGATGAT 651  
 Db 1252 GGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1311  
 Oy 652 AAGATTACGAACCCGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 711  
 Db 1312 CAAGTCAGGAACCCGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1371  
 Oy 712 GACATCAACAACTCCGCTCGACGCGGAGTGACACACAGACAC--GTGAGCGCTTC 768  
 Db 1372 GACTCAGACAACTCCGCTCGACGCGGAGTGACACAAATTAACATTAACATTAACATTA 1431  
 Oy 769 AAGCACTTCGGGAGTGGCTGAGGAGCAACAAAGCCGCGGCGGCGGCGGCGGCGGCGG 828  
 Db 1432 TCTCCGCTTGGCAGCTTGGCTCGAGACAAATGCGCGGCTATCTGACAGAAACCGGT 1491  
 Oy 829 GCGTTCATGAAACCTTGGCTCGAGCTTGGCTCGAGACAAATGCGCGGCTATCTGACAG 888  
 Db 1492 GGTGCAACGATTCAGTCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 1551  
 Oy 889 AACACGACGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 948  
 Db 1552 AACTCAGATGCTATCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 1611  
 Oy 949 ATCTGACTGAGTCCCTCGGCAAGCCGCGGCAACATGACGAGCA 994  
 Db 1612 GTCTGACGGAACACCGACTAGCTGATGATGATGATGATGATGATGATGATGATGAT 1657

RESULT 7  
 ID AA091284  
 AA091284 standard; DNA; 1849 BP.  
 XX AC  
 XX AA091284;  
 XX 13-DEC-1995 (first entry)  
 DT  
 XX T. longibrachiatum endoglucanase EGII.  
 DE  
 XX Cellulase; cellulose; signal; catalytic core; cellulase binding;  
 KW linker; ss.  
 XX Trichoderma longibrachiatum.  
 OS  
 XX Location/Qualifiers  
 FH Key  
 FT CDS  
 FT 262..324  
 FT /\*tag- a  
 FT /product= signal  
 FT /note= "seq id no 7"  
 FT 433..534  
 FT /\*tag- c  
 FT /product= linker  
 FT /note= "seq id no 23"  
 FT 535..590  
 FT /\*tag- d  
 FT /product= catalytic core domain  
 FT /note= "seq id no 15"  
 FT 591..764  
 FT /\*tag- e  
 FT Intron

FT CDS 765..1689  
 FT /tag- f  
 FT /product- catalytic core domain  
 FT /note- "seq id no 15"

XX W09516782-A.  
 XX 22-JUN-1995.  
 XX 19-DEC-1994; 94MO-US14163.  
 XX 17-DEC-1993; 93US-0169948.  
 XX (GENV ) GENENCOR INT INC.  
 XX  
 XX Clarkson KA, Collier KD, Fowler T, Larenas E, Ward M;  
 DR WPI, 1995-231574/30.  
 DR P-PSDB; AAR77264.  
 XX  
 PT Pure, truncated fungal cellulase protein from Trichoderma - useful to  
 PT reduce or eliminate dye, colourant or pigment back-staining or  
 PT redeposition in stone-washing or DIO-polishing.  
 XX  
 PS Claims 35, 58, 59; Figure 4; 105pp; English.

CC Figure 4 depicts the genomic DNA and AA sequence of EGI derived  
 CC from T. longibrachiatum. DNA fragments are claimed which comprise  
 CC SEQ ID nos 11, 13 and 15; 15 and 19; and 15, 19 and 11.  
 CC Genes for EGI and EGI1 have been isolated from T. longibrachiatum  
 CC and the protein domain structure has been confirmed. (Penttila, M.  
 CC et al., 1986, Gene, 45, 253-263; Van Arsdell, J.N. et al., 1987,  
 CC Bio/Technology, 5, 60-64; and Saloheimo, M. et al., 1988, Gene  
 CC 63, 11-21).

XX  
 XX Sequence 1849 BP; 444 A; 507 C; 440 G; 458 T; 0 other;  
 SQ

Query Match 20.9%; Score 275.6; DB 16; Length 1849;  
 Best Local Similarity 59.2%; Pred. No. 5, 1e-56;  
 Matches 489; Conservative 0; Mismatches 334; Indels 3; Gaps 1;

0Y 172 GATGGCGCGCGCAGATGAGCATTTGGCCGAGAGAGCGGCGCTCAAGCTTTTCCGATA 231  
 DB 832 GATGGCATGCGCGAGATGAGCATTTGGCCGAGAGAGCGGCGCTCAAGCTTTTCCGATA 891  
 0Y 232 TCCGCTACATGGCAGTTTCTCTCAACACAGGCTGAGCGGCGAGCTGAC 291  
 DB 892 CCTGTGCGATGGCAGTACCTGTCACACAAATTTGGGCGGCAATCTTATTCACAGAGC 951  
 0Y 292 TGGGCTCTTACACAAAGCTGTCACAGCGCTGTCGAGAGCGGCGCTTATTCATGATT 351  
 DB 952 ATTTCCAGTATGATCAGCTTTGTTCAGGGGTGCTCTCTGAGCCATATCTGATCTGTC 1011  
 0Y 352 GACATGACACACTTTGCCCGGTACACAGGCGGCGATCAGTGGTCCAGGAGCGGTGGAC 411  
 DB 1012 GACATTCACAAATTAATGCTGATGAGAGGCTGGGATCATTTGTCAGGCGGCGCTTAAAT 1071  
 0Y 412 GACATCTTTGTCAGCTCTGGGTCCGATCGCAAACTACTATGAGGACACACAGCATATC 471  
 DB 1072 GCTCAATTTACAGAGCTTTGGTGTGCGAGTTGCAATTAATGATCTGATGAGGATG 1131  
 0Y 472 ATCTTTGGCGTATGACAGCGCGGCGAGCTGATGATGATGATGATGATGATGATGATG 531  
 DB 1132 TGGTTTCGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1191  
 0Y 532 CAAAAGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 591  
 DB 1192 CAAGAGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1251  
 0Y 592 GGAACCACTTTGCGAGCTGAGAGCTATGTCACACTGGGCGGAGAGCCCTTCGGC 651  
 DB 1252 GGAATGATTTGCAATCTCTGGGCGCTTCATATGCGATGGGAGTCCAGCCGCTGTCT 1311

0Y 652 AGATTACGACCCGATGAGACACCGATTTGCTGATTTGATGTCACAGTATCTC 711  
 DB 1312 CAAGTACGACCCGATGAGATGAGTATGATGATGATGATGATGATGATGATGATGATG 1371  
 0Y 712 GACATCAACAATTCGGGTGACAGCGGCGAGTGCACACAGACAAAC--GTGAGCGCTTC 768  
 DB 1372 GACTCAGACAACTCCGGTACTACAGCGGCAATGATACAAATTAATGATGAGCGGCTTC 1431  
 0Y 769 AACGACTTCGGGAGCTGCTGAGGCGAGACACCGGCGATCATCTCCGAAGCGGC 828  
 DB 1432 TCTCCGCTTGCACCTTGGCTCCGACAGAACAAATGCGCGAGCTTCTGACAGAACCGGT 1491  
 0Y 829 GCGTCATGACCTTCGTCATGATGATGATGATGATGATGATGATGATGATGATGATG 888  
 DB 1492 GGTGCGACGTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1551  
 0Y 889 AACGAGAGCTTACATTTGGCTTTGGGCTGGGAGTGCAGGAGCTTTCACAGCTGTAC 948  
 DB 1552 AACTCAGATGCTATCTTGGCTTGGCTTGGTGGGAGTGCAGGAGCTTTCATGATGATG 1611  
 0Y 949 ATCTTACTCTGACTCCCGCGGCAAGCGGCGCACTACACGACA 994  
 DB 1612 GTCTGACGGAACACCGACTGACGCTGATGATGATGATGATGATGATGATGATGATG 1657

RESULT 8  
 AAT32223  
 ID AAT32223 standard; DNA; 1849 BP.  
 XX  
 AC AAT32223;  
 XX  
 DT 28-OCT-1996 (first entry)  
 XX  
 DE Trichoderma endoglucanase II gene.  
 XX  
 KW Endoglucanase II; EGI1; cellulase; cellulose; denim;  
 KM stonewashing; dye redeposition; backstaining; ss.  
 XX  
 OS Trichoderma longibrachiatum.  
 XX  
 FH Key Location/Qualifiers  
 FT exon 1..590  
 FT /tag- a  
 FT /codon\_start= 262..264  
 FT intron 591..764  
 FT exon /tag- b  
 FT exon 765..1849  
 FT /tag- c

XX W09623928-A1.  
 XX 08-AUG-1996.  
 XX 29-JAN-1996; 96MO-US00977;  
 XX 01-FEB-1995; 95US-0382452.  
 XX (GENV ) GENENCOR INT INC.  
 XX  
 XX Clarkson KA, Collier KD, Fowler T, Larenas E, Ward M;  
 DR WPI, 1996-371466/37.  
 DR P-PSDB; AAM02032.  
 XX  
 PT Treatment of cellulose-contg. fabrics such as denim, e.g.  
 PT stone-washing - using truncated cellulase enzyme to increase  
 PT abrasion and give reduced redeposition of dye  
 XX  
 PS Disclosure; Fig 4A-4C; 124pp; English.

CC Genomic DNA sequences (AAT32220-24) of Trichoderma longibrachiatum  
 CC respectively code for cellobiohydrolase I (CBHI) (AAM02022), CBHI  
 CC (AAM02025), endoglucanase I (EGI) (AAM02029), EGI1 (AAM02032) and EGI11





DT 13-MAR-2001 (first entry)  
 XX Trichoderma reesei EST SEQ ID NO:7427.  
 XX  
 XX Multiple gene expression: filamentous fungal cell; EST:  
 KW expressed sequence tag; Fusarium venenatum; Aspergillus niger;  
 KW Aspergillus oryzae; Trichoderma reesei; Identification; recombination;  
 KW culture condition; environmental stress; spore morphogenesis;  
 KW metabolic pathway engineering; catabolic pathway engineering; ss.  
 XX  
 XX Trichoderma reesei.  
 XX  
 XX WO200056762-A2.  
 XX  
 XX 28-SEP-2000.  
 XX  
 XX 22-MAR-2000; 2000WO-US07781.  
 XX  
 XX 22-MAR-1999; 99US-0273623.  
 XX  
 XX (NOVO ) NOVO NORDISK BIOTECH INC.  
 XX (NOVO ) NOVO NORDISK AS.  
 XX Berka RM, Rey MW, Shuster JR, Kauppinen S, Clausen IG, Olsen PB;  
 XX WPI: 2000-594572/56.  
 XX  
 XX Monitoring differential expression of genes in filamentous fungal cells  
 PT uses fluorescence-labeled nucleic acids isolated from the cells and a  
 PT substrate of expressed sequence tags -  
 XX  
 XX Claim 89; Page 2999; 316pp; English.  
 XX  
 XX The present invention describes a method for monitoring differential  
 CC expression of genes in a first filamentous fungal (FF) cell relative to  
 CC expression of the same genes in one or more second filamentous fungal  
 CC cells. The method uses fluorescence-labeled nucleic acids isolated from  
 CC the FF cells and a substrate of expressed sequence tags (EST). The ESTs  
 CC are used in the methods for monitoring differential expression of genes  
 CC in a first filamentous fungal (FF) cell relative to expression of the  
 CC same genes in one or more second filamentous fungal cells. Monitoring  
 CC the global expression of genes from FF cells allows the production  
 CC potential of the microorganisms to be improved. New genes may be  
 CC discovered, possible functions of unknown open reading frames can be  
 CC identified and gene copy number variation and stability can be  
 CC monitored. The expression of genes can be used to study how FF cells  
 CC adapt to changes in culture conditions, environmental stress, spore  
 CC morphogenesis, recombination, metabolic or catabolic pathway  
 CC engineering. Using ESTs provides several advantages over genomic or  
 CC random cDNA clones including elimination of redundancy as one spot on an  
 CC array equals one gene or open reading frame, and organisation of the  
 CC microarrays based on function of the gene products to facilitate  
 CC analysis of the results. AAF07478 to AAF11247 represents ESTs from  
 CC Fusarium venenatum; AAF11248 to AAF11853 represents ESTs from  
 CC niger; AAF11854 to AAF14878 represents ESTs from Aspergillus oryzae; and  
 CC AAF14879 to AAF15337 represents ESTs from Trichoderma reesei, which are  
 CC all specifically claimed in the present invention.  
 XX  
 XX Sequence 584 BP; 145 A; 153 C; 137 G; 149 T; 0 other;  
 S0  
 Query Match 7.7%; Score 101.4; DB 21; Length 584;  
 Best Local Similarity 59.5%; Prod. No. 1.3e-14;  
 Matches 228; Conservative 0; Mismatches 146; Indels 9; Gaps 3;

DB 129 GATTTTACGTGCGACAATACCTTGACCTCAGAACACTCCGGGTACACGCCGAATGT 188  
 OY 745 ACCACAGA-----CAACGTGACAGCGCTTCACGACTTCGGGACCTGGAGGCACAA 797  
 DB 189 ACTACAAATATACATTGACGGGCGCTTTCTCCCGCTTGCCACTTGCTCCGACACAA 248  
 OY 798 CAAGCGCCAGGCGATCATCTCCGAAGCGGCGCTGCATGGAACCTTGCTGCATGCTGC 857  
 DB 249 CAATGCCAGGCTATCTCGACAGAAACCGGTGGGCGCAAGCTTACGTCTGCATACAA 308  
 OY 858 CTTCGCGCCCAAGAACAGCCATTAGCGAAACAGCAGCTTCATATGCTTTGGG 917  
 DB 309 CATGTGCCAGCAAAATCCATATCTCAACACAGACTCATATGCTATGTTGG 368  
 OY 918 CTGGGTCGCGGACGCTTGACA 940  
 DB 369 TTGGGGTCCGATCATTTGATA 391

## RESULT 12

.....AAV16436  
 ID AAV16436 standard; DNA; 912 BP.

XX AAV16436;

DT 18-AUG-1998 (first entry)

XX Glucanase II gene.

XX Endo-beta-1,4-glucanase II; transgenic; degradation; glucan; ss.

XX Aspergillus niger.

OS Aspergillus niger.

XX Key Location/Qualifiers

FT CDS 1..912

FT /tag- a

FT /product- "Endo-beta-1,4-glucanase II enzyme"

FT /transl\_except- (pos:16..18, aa:Trp)

FT /transl\_except- (pos:289..291, aa:Tyr)

FT /transl\_except- (pos:388..390, aa:Asn)

FT /note- "No stop codon given"

XX W09806858-A1.

XX 19-FEB-1998.

XX 11-AUG-1997; 97WO-EP04415.

XX 15-AUG-1996; 96GB-0017184.

XX (DANI-) DANISCO AS.

XX Gravesen TN, Rasmussen P;

XX WPI: 1998-159548/14.

XX P-PSDB; AAW46807.

XX New Aspergillus niger beta-1,4-endo:glucanase - used for the

PT production of foodstuffs, feed and detergents and in brewing, pulp

XX and textile industries

XX Claim 6; Page 58; 89pp; English.

XX The glucanase II gene encodes the endo-beta-1,4-glucanase II enzyme.

CC The protein can be expressed in filamentous fungi, plant and yeast

CC systems. The enzyme can be used to degrade glucans. This can be used

CC for the preparation of foodstuffs, including feeds, as well as in the

CC brewing, pulp and textile industries and in the formulation of

CC detergents.

XX Sequence 912 BP; 224 A; 231 C; 268 G; 189 T; 0 other;

Query Match 6.0%; Score 79.4; DB 19; Length 912;  
 Best Local Similarity 47.5%; Pred. No. 2.6e-09;  
 Matches 365; Conservative 0; Mismatches 371; Indels 33; Gaps 3;

QY 205 GAGACGGCCCTCAGCTCTTTCGATATCCGCTACATGGCAGTTTGTCTCAACACAG 264  
 DB 121 GACAAAGGGATGAACTTCTCCGCTCCAGTTCATGATGAGAGTTGTCTGCCACTCG 180  
 QY 265 GTGGAGCGCAAGCTGACGAGCTCACTGGGCTCTACAGCAAGGTCGACGGCTGT 324  
 DB 181 ATGACTGCTTCATATATATGAGATATCTGGCCACTTGCAGCAGATGATAAAGCGTA 240  
 QY 325 CTCGACAGCGCGCCCTACTGATGATGATGATGATGATGATGATGATGATGATGAT 384  
 DB 241 AGGAGAGGAGCGCCCATGGCTGTGTCAGACCCCTAATAGATGATGATGATGATGATG 300  
 QY 385 ATCATGCGCCAGGAGCGCTGTGTCAGACATCTTGTGATGATGATGATGATGATGATG 444  
 DB 301 ATCATCTCCAGCAGCTGACCTTCCAGACCTTCTGGAGAACCTGGCGGCGCAG----- 354  
 QY 445 AAGTACTGAGAGAGCAAGCAAGATCATCTTGGCTGATGAAGAGCGCGCAGACCTC 504  
 DB 355 -----TACAAAGATTAAGAGCTGCTGATGATGATGATGATGATGATGATGATG 408  
 QY 505 GACATTGAGATCTGGCGCAGACGTCGCAAAAGTGTGATGATGATGATGATGATGATG 564  
 DB 409 GACCAAGATCTGCTGTAACCTCAACAGCAGCCATTAAGGCGATCCGCGCGCAGGT 468  
 QY 565 GCCACTCGCAGATGATCTCTGCTGCGCCGACCACTTTCGACGCTGACGATGTG 624  
 DB 469 GCGACGAGCAGTACATCTTCTGTCGAAGGACACTCTGAGACCGCGCTGACGCTGGT 528  
 QY 625 TCCACTGGAGCGGAGAGCGCTCGGCAAGATTAGCAACCGGATGAGACCGACGATT 684  
 DB 529 -----GACGTCAAGCAGACATGAAAGATTGACCGACCGCGAAGACAG 573  
 QY 685 CTGTACTTGTGATGTCAGATGATGATGATGATGATGATGATGATGATGATGATG 744  
 DB 574 ATGCTTAATGAATGACAGCAGTACGATGATGATGATGATGATGATGATGATGATG 633  
 QY 745 ACCACAGACAGCTC-----GACGCTTCACAGCTTCCGCGAGCTGTGAGGACAG 788  
 DB 634 GTGTCCGAGACCTCGGAAAGAGCGGCTGCTGACGAGCTGACAGCTGCTGAGAGCAAT 693  
 QY 799 AAGCGCAGGCGCATCTCTCCGAAAGGGGCGCTCATGGAACCTTCGTGATGATGCTGC 858  
 DB 694 AAGAGGCTGCTTCAATAGCGGAATATGCCGGGCTTCATGATGATGATGATGATGATG 753  
 QY 859 TTCTGCGCCAGAAAGAGCCATTAGCGAAAGACGAGCTTACATTTGGCTTTGGGC 918  
 DB 754 GTGTGCGGATGCTGAGATGATGATGATGATGATGATGATGATGATGATGATGATG 813  
 QY 919 TGGGGTCCGCGAGCTTTGACAGCTGATGATGATGATGATGATGATGATGATGATG 967  
 DB 814 TGGGAGACCGGCGCATGTGAGGAGAGTACATTTTTCAGCATGAGAGCC 862

RESULT 13  
 AAD20928  
 ID AAD20928 standard; DNA; 1008 BP.  
 XX  
 AC AAD20928;  
 XX  
 DT 15-JAN-2002 (first entry)  
 XX  
 DE Talaromyces emersonii beta-glucanase CEA DNA.  
 XX  
 KW Beta-glucanase; CEA; antileptic; fungicide; hyperlipaemia; animal feed;  
 KM fabric; edible foodstuff; textile; brewing; distilling; bioremediation;  
 KW dental hygiene; leather-treatment; paper manufacture; plant extract;  
 KM detergent treatment; baking; ds.  
 XX  
 OS Talaromyces emersonii.

XX  
 FH Key Location/Qualifiers  
 FT CDS 1.1008  
 FT /tag 2  
 FT /product "T. emersonii CEA protein"  
 PN W0200170998-A1.  
 XX  
 PD 27-SEP-2001.  
 XX  
 PF 20-MAR-2001; 2001WO-EP03174.  
 XX  
 PR 20-MAR-2000; 2000BP-0302263.  
 XX  
 PA (STAM ) DSM NV.  
 XX  
 PI Van Den Hombergh JPTW, Van Der Laan J, Daran JG, Herweijer MA;  
 PI Teufel DP;  
 DR MPI: 2001-648392/74.  
 DR P-PSDB; AAE12786.  
 PT New polypeptide for treating hyperlipemia and/or high serum  
 PT cholesterol and triglyceride levels, comprises the beta-glucanase  
 PT protein obtainable from Talaromyces fungus  
 PS Claim 5; page 64-65; 76pp; English.  
 XX  
 CC The invention relates to a beta-glucanase polypeptide obtainable from a  
 CC fungus of the genus Talaromyces, e.g. T. emersonii, having endoglucanase  
 CC activity (EC 3.2.1.4). Beta-glucanase polypeptide is useful for the  
 CC manufacture of a medicament for treating hyperlipaemia, high serum  
 CC cholesterol and triglyceride levels. Beta-glucanase polypeptide are  
 CC useful for treating fungal or plant materials (plant pulp, plant  
 CC extracts), edible foodstuffs or ingredients, or fabrics, textiles or  
 CC clothes containing plant materials. Beta-glucanase polypeptide is  
 CC useful for reducing the viscosity of a plant material, for cleaving  
 CC beta-D-glucan polymers in the plant material, and for processing plant  
 CC pulp, juice or extract, by incubating the pulp, juice or extract.  
 CC Beta-glucanase polypeptide is applicable in brewing, distilling,  
 CC bioremediation, dental hygiene, leather-treatment, paper manufacture,  
 CC textile treatment or manufacture, baking or bread making, washing or  
 CC detergent treatment, treating flower bulbs or in animal feed.  
 CC Beta-glucanase polypeptide is also useful during the production of milk  
 CC substituted from soybean. The present sequence is Talaromyces emersonii  
 CC beta-glucanase CEA DNA.  
 CC  
 XX  
 SQ Sequence 1008 BP; 220 A; 298 C; 304 G; 186 T; 0 other;

Query Match 6.0%; Score 79; DB 22; Length 1008;  
 Best Local Similarity 47.8%; Pred. No. 3.3e-09;  
 Matches 355; Conservative 0; Mismatches 355; Indels 33; Gaps 3;

QY 193 CATTTGCGCGAAGCAGCGCTCAAGCTTTCGATATCCGCTACATGGCAGATTGTC 252  
 DB 199 CAGATTCATTCAGCAGCGGAGATGATGATGATGATGATGATGATGATGATGATG 258  
 QY 253 CTCACACACAGGTCAGCGCAAGCTGACGAGCTCACTGAGGCTCTCAACAAAGTTC 312  
 DB 259 GTGCCCAACCAAGATGAGCGGCGGCTGATTCGCTATTTCCAGAGGCTACAGCCAGGTT 318  
 QY 313 GTCAAGCGCTGTCTCGAGAGCGGCGCTACTGATGATGATGATGATGATGATGATGATG 372  
 DB 319 ATCAACTATATTAACAGCCATGAGCGCTGCGCATGATTAACCGCATTAAGTGGCGGA 378  
 QY 373 TACAACGCGCGCATCATCGCCAGGAGCGGTGCGAGCAGCATTTTGTGACACTCTG 432  
 DB 379 TACTACAGCAATATCATCTCC-----TCGCGCTGTGACTTCCAGACTTCTGG 426  
 QY 433 GTCCAGATCCCAAGTACTGAGAGCAGACAGCAAGTATCTTTGGCTGTGATGACGAG 492  
 DB 427 CACACTATTCGCTCCCACTTTGGGATTAATGACAAATGCTATTTTCGACAGCAACAGGA 486

OY 493 CCGACGACCTTCGACATTTAGATCTGGGCGCAGACGTCGCAAAAGGCTGCTACCTGACATC 552  
 DB 487 TACACGACATGAGCAAGAAAGCTTGTCTCCAGCTCAACGAGCCGCTGACGAGCATC 546  
 OY 553 CGAAAGGCGGCGGACCTGCGAGATGATCTCTGCGCCGGAACCAATTGGCCAGCCTC 612  
 DB 547 CGCGCGGCGGCGGACATCAACAGTACATCTTCTGCGAGGCACTCGTGACCGGGCC 606  
 OY 613 GAGACGATGTGTCCACTGCGACGCGGAGACCCCTGCGGCAAGATTACGACCCGATGGA 672  
 DB 607 TGGACAT-----GGACGCGAGGTCAACGACGAGGATGGGCAACCTGACGAGAC 651  
 OY 673 AGCACCATTGCTGTACTTTGATGTCCACAAGTATCTGCAATCACTACACTCCGGGTG 732  
 DB 652 CCGGAGAAACAAGATCTGTACGAGATGACACGATCTGAGACGCGGTGGGCAAGC 711  
 OY 733 CAGCCGAGTGCACACAGACACAGTC-----GACGCGCTTCAACGACTTCCGGGACAG 786  
 DB 712 TCGGACCACTGCTCACTGACATCGGCGGAGACGCGCTGAGTCCGCGACGCGCTG 771  
 OY 787 CTGAGGCGAAGCAAGCGCCAGCCATCTCTCGAAGGCGGCGGTCCATGGAACCTTCG 846  
 DB 772 CTGAAGCGAAGCGCAAGAAAGCGGATCTGCGGCGGAGTACGCTGGGCGGCGCAACGCGTG 831  
 OY 847 TGCAATGACTGCTTCTGTGCGCCAGAACAGCCATTAGCGAAACAGCGAGCTTACTT 906  
 DB 832 TGCAGAGAGCGCGCTACCGGCGATCTGCACTATCTGCGCAACATATGATGTGTGACC 891  
 OY 907 GGCCTTGTGGGCTGGGGTGGCGG 929  
 DB 892 GGTCTATCTGTGTGGCGGCTGG 914

RESULT 14  
 ABQ44816/c  
 ID ABQ44816 standard; DNA; 985 BP.

XX ABQ44816:

DT 13-JUL-2002 (first entry)

DE oligonucleotide for detecting cytosine methylation SEQ ID NO 31407.

KW Human: cytosine methylation; 5'-CPG-3'; uracil: cytosine; diagnosis;  
 KW drug; side effect; cancer; central nervous system; cardiovascular;  
 KW gastrointestinal; respiratory system; single nucleotide polymorphism;  
 KW SNP; cell differentiation; ds.

XX Homo sapiens.

OS W0200218632-A2.

PD 07-MAR-2002.

PF 01-SEP-2001; 2001WO-EPI10074.

PR 01-SEP-2000; 2000DE-1043826.

PR 05-SEP-2000; 2000DE-1044543.

PA (EPiG-) EPIGENOMICS AG.

PI Olek A, Piepenbrock C, Berlin K, Guetig D;

DR WPI; 2002-371829/40.

PT Determining the degree of cytosine methylation in genomic DNA, useful

PT for diagnosis and prognosis, comprises selective hybridization of

PT amplicons from chemically treated DNA

PS Claim 12: 56bp + Sequence Listing: 56bp; Germany;  
 CC This invention describes a novel method for determining the degree of  
 CC methylation of a particular cytosine in a motif 5'-CPG-3', present in a

CC genomic sample of DNA. The sample is treated chemically to convert  
 CC cytosine (C) but not methylated C, to uracil, then part of the genomic  
 CC DNA that contains the target C is amplified to form a labeled amplicon.  
 CC The amplicon is hybridized to two classes, each with at least one  
 CC member, of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers  
 CC and the degree of hybridization to both classes is determined from the  
 CC label on the amplicon. From the ratio of labels hybridized to the two  
 CC classes of oligomers, the degree of methylation is calculated. The method  
 CC is used: (1) for diagnosis and/or prognosis of side effects of  
 CC therapeutic drugs and of a wide range of diseases, e.g. cancer, disorders  
 CC of the central nervous, cardiovascular, gastrointestinal and respiratory  
 CC systems etc., particularly by detecting mutations or single nucleotide  
 CC polymorphisms (SNP's); and (11) for differentiation of cell or tissue  
 CC types and for investigating cell differentiation. The method allows the  
 CC methylation status of many C residues to be determined simultaneously.  
 CC AB013410-AB054121 represent genomic DNA sequences used to illustrate the  
 CC method for determining the degree of cytosine methylation described in  
 CC the disclosure of the invention.

XX Sequence 985 BP; 234 A; 145 C; 326 G; 280 T; 0-other;

Query Match 5.6%; Score 74.2; DB 24; Length 985;

Best Local Similarity 46.7%; Pred. No. 4.6e-08; Matches 342; Conservative 0; Mismatches 383; Indels 8; Gaps 3;

OY 51 CTTGCGCGGCAAGATCAATATCTGGGCGTCCATTCGCGAATCGACTTTGGCTGGA 110  
 DB 840 CGTGCAGCTGACATCGACATGCGATGCGATGCGAGCTCGACGCGCATCGACGTGA 781  
 OY 111 CATGACGCGGAGCTGTGCGACTGACAGCTGTGTGCGCCCTGCGTCAACAAAGG 170  
 DB 780 GGTCTACATTCGAC--GTGACATGACATCTACATCGATCTACATCGATCAACGTGACGT 723  
 OY 171 AGATGGCGCGCGCAGATGATGAGCATTTGCGCCAGACGCGCGCTCAACGCTTTTCGAT 230  
 DB 722 CTACATGACGTCGAACGTCACATCTACATGACATCAACATCTACATCTACATCTACGT 663  
 OY 231 ATCCGCTACATGCGAGTTGTCTCAACACACGCGTGCGCAAGCTGACGAGCTCAA 290  
 DB 662 CGACATCGACCTCGACATCGACATCCGCTTACATCTACATCTACATCTACATCTACGT 603  
 OY 291 CTGGGCGCTCCACAAAGGTCGTCACAGCGCTGTCTGAGACGCGGCGCTTCTGATGAT 350  
 DB 602 CTACATCGACCTCGACATCTACATCTACATCTACATCTACATCTACATCTACATCTACGT 543  
 OY 351 TGACATGACAACTTTGCGCCGCTACAGCGCGGATCTACGCGCAAGGAGCGTGTGGA 410  
 DB 542 CGAGCTTACATCTACATCTACATCTACATCTACATCTACATCTACATCTACATCTACAT 483  
 OY 411 CGACATCTTTGCGACTCTGGGTCCAGAT---CGCAAGTACTTACGAGCAACAGCAA 467  
 DB 482 CTACATCTACATCTACATCTACATCTACATCTACATCTACATCTACATCTACATCTACGT 423  
 OY 468 GATCATCTTTGGCGTGAAGACGAGCGCGACGCTGACATCTACATCTACATCTACATCTACGT 527  
 DB 422 CTACATCTACATCTACATCTACATCTACATCTACATCTACATCTACATCTACATCTACGT 363  
 OY 528 GTGCCAAAGGTCGTACTGCGATCCGAAGCGCGCGCCACTGCGAGATGATCTTCT 587  
 DB 362 CTAGCTGACATCTACATCTACATCTACATCTACATCTACATCTACATCTACATCTACAT 303  
 OY 588 GCCCGGAACCACTTTGCGAGCTGAGAGCTATGTCTCACTGGCGACCGGGAAGCCCT 647  
 DB 302 CTAGCTTACATCTACATCTACATCTACATCTACATCTACATCTACATCTACATCTACGT 243  
 OY 648 CGGCAAGATTACGAACCGGGA---TGAAGACCGGATTTGCTGATCTTGTATGCCA 704  
 DB 242 CGACTTCACATCAACGTCGACATCGACATCAATCTACATCTACATCTACATCTACAT 183  
 OY 705 GTATCTGACATCAACATCTCGGCGTCCAGCGAGTGCACCAACAGCAACGTCGAGCG 764  
 DB 182 CTAGCTTACATCAACATCTACATCTACATCTACATCTACATCTACATCTACATCTACGT 123

Qy 765 CTTCAACGACTTC 777  
 Db 122 CTACATCAACATC 110

## RESULT 15

AB044817  
 ID AB044817 standard; DNA; 985 BP.

AC AB044817;

DT 12-JUL-2002 (first entry)

DE Oligonucleotide for detecting cytosine methylation SEQ ID NO 31408.

Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;  
 drug; side effect; cancer; central nervous system; cardiovascular;  
 gastrointestinal; respiratory system; single nucleotide polymorphism;  
 SNP; cell differentiation; ds.

OS Homo sapiens.

PN WO200218632-A2.

PD 07-MAR-2002.

PF 01-SEP-2001; 2001WO-EP10074.

PR 01-SEP-2000; 2000DE-1043826.

PR 05-SEP-2000; 2000DE-1044543.

PA (EPIG-) EPIGENOMICS AG.

PI Olek A, Plepenbrock C, Berlin K, Guetig D;

DR WPI; 2002-371829/40.

PT Determining the degree of cytosine methylation in genomic DNA, useful  
 for diagnosis and prognosis, comprises selective hybridization of  
 PT amplicons from chemically treated DNA

PS Claim 12; 56pp + Sequence Listing; 56pp; German.

CC This invention describes a novel method for determining the degree of  
 CC methylation of a particular cytosine in a motif 5'-CpG-3', present in a  
 CC genomic sample of DNA. The sample is treated chemically to convert  
 CC cytosine (C) but not methylated C, to uracil, then part of the genomic  
 CC DNA that contains the target C is amplified to form a labeled amplicon.  
 CC The amplicon is hybridised to two classes, each with at least one  
 CC member, of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers  
 CC and the degree of hybridisation to both classes is determined from the  
 CC label on the amplicon. From the ratio of labels hybridised to the two  
 CC classes of oligomers, the degree of methylation is calculated. The method  
 CC is used: (i) for diagnosis and/or prognosis of side effects of  
 CC therapeutic drugs and of a wide range of diseases, e.g. cancer, disorders  
 CC of the central nervous, cardiovascular, gastrointestinal and respiratory  
 CC systems etc., particularly by detecting mutations or single nucleotide  
 CC polymorphisms (SNP's); and (ii) for differentiation of cell or tissue  
 CC types and for investigating cell differentiation. The method allows the  
 CC methylation status of many C residues to be determined simultaneously.  
 CC AB013410-AB051421 represent genomic DNA sequences used to illustrate the  
 CC method for determining the degree of cytosine methylation described in  
 CC the disclosure of the invention.

CC Sequence 985 BP; 280 A; 326 C; 145 G; 234 T; 0 other;

Query Match 5.6%; Score 74.2; DB 24; Length 985;

Best Local Similarity 46.7%; Pred. No. 4.6e-08;

Matches 342; Conservative 0; Mismatches 383; Indels 8; Gaps 3;

Qy 51 CCTCCGCGGCAAGATCAATATCTGGGCTGGCCATTCGCCGAGATCGACTTGGCTGCGA 110  
 Db 146 CGTCGACGTCGATCGACATCGATCGATCGATCGATCGATCGATCGATCGATCGATCGA 205

Qy 111 CATCGACGGCAGCTGTCCGACTGTACAGCTGTGTGTGCCCCCTGCTGAGTACAAAGAGG 170  
 Db 206 CGTCTACATCGAC--GTGACATCGACATCTACATGACATCTACATGACATGACATGACAT 263  
 Qy 171 AGATGGCGCCGGCCAGATGAAGCATTTGCCGAAGACGAGCGGCTTCATTTGGCAT 230  
 Db 264 CTACATCGACGTCGATCAACATCTACATCGACATCAACATCTACATCTACATCTACAT 323  
 Qy 231 ATCCGCTACATGGCAGTTTGTCTTCAACAACAGGCTGAGCGGACCTGAGAGCTCAA 290  
 Db 324 CGACATCGACGTCGACATGACATCGCGCTTACATCTACATCTACATCTACATCTACAT 383  
 Qy 291 CTGGGGCTCTCTACCAAGAGTGTCAAGCGCTGTGAGAGCGGGCCCTACTGATGAT 350  
 Db 384 CTACATCGACGTCGACATCTACATCTACATCTACATCTACATCTACATCTACATCTAC 443  
 Qy 351 TGACATGACAACTTTGCCGCTACACAGCGGCGCATCTGCGGAGGAGGCGTTCGGA 410  
 Db 444 CGACGTCATCATCTACGTCGAGCTGACGCTTACATCTACATCTACATCTACATCTACAT 503  
 Qy 411 CGACATCTTTTCGACCTCTGGTCCAGAT--CGCAAGTACTACGAGACACGACAA 467  
 Db 504 CTACATCTACATCTACGTCGAGCTTACATCTACGTCGACATCTACATCTACATCTAC 563  
 Qy 468 GATCATCTTTGGCTGTATGAGAGCGGACGACGCTGACATCTGAGATCTGGGCGGAGAC 527  
 Db 564 CTACATCTACGTCATCTACATCTACATCTACATCTACATCTACATCTACATCTACAT 623  
 Qy 528 GTGCCAAAGAGTGTGCTACATGCTGCAAGAGCGGCGGCGGCGGCGGCGGCGGCGG 587  
 Db 624 CTACGTCGACATGACATGACATGACATGACATGACATGACATGACATGACATGACAT 683  
 Qy 588 GCCCGGACCACTTTGCCAGCGTGTGACAGCTGTGTCATCTGCTGAGCGGAGACCT 647  
 Db 684 CTACGTCATCTACATCTACGTCGACATCTACATCTACATCTACATCTACATCTACAT 743  
 Qy 648 CGGCAAGATTACGACCCGGA--TGAAGACCGATTTGCTGTACTTTGATGATGACAA 704  
 Db 744 CGACATCGACATCAACGTCGACATCAACATCTACATCTACATCTACATCTACATCTAC 803  
 Qy 705 GATCTCGACATCAACACTCGGCTGCGACGCGGAGTGCACCCAGACAGACATCGACGC 764  
 Db 804 CTACGTCATCAACATCAACATCTACATCTACATCTACATCTACATCTACATCTACAT 863  
 Qy 765 CTTCAACGACTTC 777  
 Db 864 CTACATCAACATC 876

Search completed: May 15, 2003, 16:17:24  
 Job time : 289.624 secs



GenCore version 5.1.4 p5.4578  
Copyright (c) 1993 - 2003 Comugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 15, 2003, 02:49:23 ; Search time 50.7022 Seconds  
(without alignments)  
7965.992 Million cell updates/sec

Title: US-10-028-245-4

Perfect score: 1317

Sequence: 1 atgcgcgcacactccctctc.....acacgcctgtagcgttttga 1317

Scoring table: IDENTITY\_NUC

Gapop 10.0, Gapext 1.0

Searched: 441362 seqs, 15338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database: Issued\_Patents\_NA.\*

1: /cgn2.6/ptodata/1/ina/5A.COMB.seq:\*  
2: /cgn2.6/ptodata/1/ina/5B.COMB.seq:\*  
3: /cgn2.6/ptodata/1/ina/6A.COMB.seq:\*  
4: /cgn2.6/ptodata/1/ina/6B.COMB.seq:\*  
5: /cgn2.6/ptodata/1/ina/PCUTUS.COMB.seq:\*  
6: /cgn2.6/ptodata/1/ina/backfiles1.seq:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	289.8	22.0	1463	US-09-254-733-6	Sequence 6, Appl
2	275.6	20.9	1155	US-08-169-948B-15	Sequence 15, Appl
3	275.6	20.9	1155	US-08-448-873-15	Sequence 15, Appl
4	275.6	20.9	1155	US-08-382-452D-15	Sequence 15, Appl
5	55.6	4.2	390	US-09-197-649-7	Sequence 7, Appl
6	53.8	4.1	1241	US-08-471-033-39	Sequence 39, Appl
7	53.8	4.1	1241	US-08-471-033-42	Sequence 42, Appl
8	53.8	4.1	1241	US-08-471-044-39	Sequence 39, Appl
9	53.8	4.1	1241	US-08-471-044-42	Sequence 42, Appl
10	53.8	4.1	1241	US-08-463-483A-39	Sequence 39, Appl
11	53.8	4.1	1241	US-08-463-483A-42	Sequence 42, Appl
12	53.8	4.1	1241	US-08-471-046A-39	Sequence 39, Appl
13	53.8	4.1	1241	US-08-471-046A-42	Sequence 42, Appl
14	53.8	4.1	1241	US-08-470-566B-39	Sequence 39, Appl
15	53.8	4.1	1241	US-08-470-566B-42	Sequence 42, Appl
16	53.8	4.1	1241	US-08-469-334-39	Sequence 39, Appl
17	53.8	4.1	1241	US-08-469-334-42	Sequence 42, Appl
18	53.8	4.1	1241	US-09-300-529-39	Sequence 39, Appl
19	53.8	4.1	1241	US-09-300-529-42	Sequence 42, Appl
20	53.8	4.1	1358	US-08-471-033-45	Sequence 45, Appl
21	53.8	4.1	1358	US-08-471-044-45	Sequence 45, Appl
22	53.8	4.1	1358	US-08-463-483A-45	Sequence 45, Appl
23	53.8	4.1	1358	US-08-471-046A-45	Sequence 45, Appl
24	53.8	4.1	1358	US-08-470-566B-45	Sequence 45, Appl
25	53.8	4.1	1358	US-08-469-334-45	Sequence 45, Appl
26	53.8	4.1	1358	US-09-300-529-45	Sequence 45, Appl
27	53.8	4.1	1389	US-08-471-033-27	Sequence 27, Appl

28	53.8	4.1	1389	US-08-471-044-27	Sequence 27, Appl
29	53.8	4.1	1389	US-08-463-483A-27	Sequence 27, Appl
30	53.8	4.1	1389	US-08-471-046A-27	Sequence 27, Appl
31	53.8	4.1	1389	US-08-470-566B-27	Sequence 27, Appl
32	53.8	4.1	1389	US-08-469-334-27	Sequence 27, Appl
33	53.8	4.1	1389	US-09-300-529-27	Sequence 27, Appl
34	53.8	4.1	1399	US-08-471-033-24	Sequence 24, Appl
35	53.8	4.1	1399	US-08-471-044-24	Sequence 24, Appl
36	53.8	4.1	1399	US-08-463-483A-24	Sequence 24, Appl
37	53.8	4.1	1399	US-08-471-046A-24	Sequence 24, Appl
38	53.8	4.1	1399	US-08-470-566B-24	Sequence 24, Appl
39	53.8	4.1	1399	US-08-469-334-24	Sequence 24, Appl
40	53.8	4.1	1399	US-09-300-529-24	Sequence 24, Appl
41	53.8	4.1	4031	US-08-471-033-49	Sequence 49, Appl
42	53.8	4.1	4031	US-08-471-044-49	Sequence 49, Appl
43	53.8	4.1	4031	US-08-463-483A-49	Sequence 49, Appl
44	53.8	4.1	4031	US-08-471-046A-49	Sequence 49, Appl
45	53.8	4.1	4031	US-08-470-566B-49	Sequence 49, Appl

## ALIGNMENTS

RESULT 1  
US-09-254-733-6  
Sequence 6, Application US/09254733  
Patent No. 6277596  
GENERAL INFORMATION:  
APPLICANT: WATANABE, MANABU  
APPLICANT: MORIYA, TATSUKI  
APPLICANT: AOYAGI, KAORU  
APPLICANT: SUMIDA, NAOMI  
APPLICANT: MURAKAMI, TAKESHI  
TITLE OF INVENTION: REGULATORY SEQUENCE OF CELLULOSE CBH1 GENES ORIGINATING  
TITLE OF INVENTION: IN TRICHODERMA VIRIDE AND SYSTEM FOR MASS-PRODUCING  
FILE REFERENCE: 99-0266\*/LC(WMC)/00144  
CURRENT APPLICATION NUMBER: US/09/254,733  
NUMBER OF SEQ ID NOS: 52  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 6  
LENGTH: 1463  
TYPE: DNA  
ORGANISM: TRICHODERMA VIRIDE MC300-1  
FEATURE:  
NAME/KEY: sig\_peptide  
LOCATION: (14)..(76)  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (14)..(76)  
FEATURE:  
NAME/KEY: mat\_peptide  
LOCATION: (77)..(1450)  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (77)..(342)  
FEATURE:  
NAME/KEY: Intron  
LOCATION: (343)..(525)  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (526)..(1450)  
US-09-254-733-6  
Query Match 22.0%; Score 289.8; DB 4; Length 1463;  
Best local similarity 59.8%; Pred. No. 1.8e+58;  
Matches 505; Conservative 0; Mismatches 337; Indels 3; Gaps 1;  
QY 172 GATGCGCGCGCCAGATGAGCATTTTCCGCGAAGCAGCGGCTCAAGCTCTTTGCCATA 231  
DB 593 GACGGATATGCGCCAGATGAGCATTTTCCGCGAAGCAGATGAGCTATTTTCCGCGATA 652

QY 232 TCCGCTACATGCGAGTTTGTCTCTCAACACAGGTCGAGCGTGGACAGCTGACAGCTCAAC 291  
 Db 653 CCGGTCGATGTCAGTACTGCTTAAACACATCTGGGTGAGACTCTCGATTCCACCACT 712  
 QY 232 TGGGCTCTTACACAGGTCGTCAACGCTGTCTCGAGACGGGGCCCTACTGATGAT 351  
 Db 713 ATCTGGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 772  
 QY 332 GACATGCAAACTTGGCCCGCTACACAGGCGGCTATCGGCTGAGGAGGCGCTGTGGAC 411  
 Db 773 GACATGCAAACTTGGCCCGCTACACAGGCGGCTATCGGCTGAGGAGGCGCTGTGGAC 832  
 QY 412 GACATGCTTTGTCGACCTGTGGGTCGAGATCGCAAGTACTGAGAGCAACAGCAAGATC 471  
 Db 833 GCCCGATTACACAGCTTTGTGGTCGAGTGGGATGAGTACGCTCTCAAGTACAGGCTG 892  
 QY 472 ATCTTGGCTGTGAGACGAGCGGACGACCTGACATTTGAGATGATGATGATGATGATGAT 531  
 Db 893 TGGTTCGGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 952  
 QY 532 CAAAGGTCGTCAGTGCATCGCAAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 591  
 Db 953 CAAAGGTCGTCAGTGCATCGCAAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1012  
 QY 592 GGAACCAACTTGGCGAGGTCGAGAGCTATGTCACCTGGCAGCGCGGGAAGCGGCGGCGG 651  
 Db 1013 GGAATGATTTATCAATCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1072  
 QY 652 AAGATTACGAACCGGATGGAAGCAGGATTTGCTGATCTTGTGATGATGATGATGATGATGAT 711  
 Db 1073 CAGGTAACGAACCGGATGGAAGCAGGATTTGCTGATCTTGTGATGATGATGATGATGATGAT 1132  
 QY 712 GACATCAACACTCGGCGTCCGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 768  
 Db 1133 GACTGCGCAACTCGGCGTCCGAGCTACGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1192  
 QY 769 AACGACTTGGCGAGCTGCTGAGGAGCAAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 828  
 Db 1193 GCTGCTCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1252  
 QY 829 GCGTCATGGAACCTTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 888  
 Db 1253 GGTGGCAATGTTCACTGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1312  
 QY 889 AACGACGCTCTACATTTGGCTTTGTGGGCTGGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 948  
 Db 1313 AACGACGCTCTACATTTGGCTTTGTGGGCTGGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1372  
 QY 949 ATCTTGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1008  
 Db 1373 ATCTTGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1432  
 QY 1009 GAGTG 1013  
 Db 1433 TCGTG 1437

CITY: South San Francisco  
 STATE: CA  
 COUNTRY: USA  
 ZIP: 94080  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Patentin Release #1.0, Version #1.25  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/169,948B  
 FILING DATE: DEC 17 1993  
 CLASSIFICATION: 435  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Horn, Margaret A.  
 REGISTRATION NUMBER: 33,401  
 REFERENCE/DOCKET NUMBER: GC226  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (415) 742-7536  
 TELEFAX: (415) 742-7217  
 INFORMATION FOR SEQ ID NO: 15:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 1155 base pairs  
 TYPE: nucleic acid  
 STRANDEDNESS: single  
 TOPOLOGY: linear  
 MOLECULE TYPE: DNA (genomic)  
 FEATURE:  
 NAME/KEY: CDS  
 LOCATION: join(1..56, 231..1155)  
 US-08-169-948B-15

Query Match 20.9%; Score 275.6; DB 2; Length 1155;  
 Best Local Similarity 59.2%; Pred. No. 3.3e-55;  
 Matches 489; Conservative 0; Mismatches 334; Indels 3; Gaps 1;

QY 172 GATGCGCGCGCGCGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 231  
 Db 298 GATGCGCGCGCGCGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 357  
 QY 232 TCCGCTACATGCGAGTTTGTCTCTCAACACAGGTCGAGCGTGGACAGCTGACAGCTCAAC 291  
 Db 357 GATGCGCGCGCGCGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 417  
 QY 358 CCGTGCAGATGCGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 477  
 Db 477 TGGGCTCTTACACAGGTCGTCAACGCTGTCTCGAGACGGGCGGCTACTGATGAT 351  
 QY 418 ATTTCAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 477  
 Db 478 GACATGCAAACTTGGCCCGCTACACAGGCGGCTATCGGCTGAGGAGGCGCTGTGAT 537  
 QY 532 GACATGCAAACTTGGCCCGCTACACAGGCGGCTATCGGCTGAGGAGGCGCTGTGAT 411  
 Db 537 GACATGCAAACTTGGCCCGCTACACAGGCGGCTATCGGCTGAGGAGGCGCTGTGAT 537  
 QY 412 GACATGCTTTGTCGACCTGTGGGTCGAGATCGCAAGTACTGAGAGCAACAGCAAGATC 471  
 Db 538 GCTCAATTCAGAGGCTTTGTGGTCGAGTGGCATCAAGTACGATCTCACTGAGGATG 597  
 QY 472 ATCTTGGCTGTGAGACGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 531  
 Db 598 TGGTTCGCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 657  
 QY 532 CAAAGGTCGTCAGTGCATCGCAAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 591  
 Db 658 CAAAGGTCGTCAGTGCATCGCAAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 717  
 QY 592 GGAACCAACTTGGCGAGGTCGAGAGCTATGTCACCTGGCAGCGGGAAGCGGCGGCGGCGGCGG 651  
 Db 718 GGAATGATTTGCAATCTGCTGGGCTTTTATATTCGATGCGATGCGATGAGCGCGGCTGTCT 777  
 QY 652 AAGATTACGAACCGGATGGAAGCAGGATTTGCTGATCTTGTGATGATGATGATGATGATGATGAT 711  
 Db 778 CAAAGTACGAACCGGATGGAAGCAGGATTTTGTGATGATGATGATGATGATGATGATGATGAT 837  
 QY 712 GACATCAACACTCGGCGTCCGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 768

RESULT 2  
 US-08-169-948B-15  
 Sequence 15, Application US/08169948B  
 Patent No. 5861271  
 GENERAL INFORMATION:  
 APPLICANT: Fowler, Timothy  
 APPLICANT: Ward, Michael  
 APPLICANT: Clarkson, Kathleen  
 APPLICANT: Collier, Katherine  
 APPLICANT: Larens, Edmund  
 TITLE OF INVENTION: No. 5861271 Cellulase Enzymes and Systems  
 TITLE OF INVENTION: For Their Expression  
 NUMBER OF SEQUENCES: 48  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Genencor International  
 STREET: 180 Kimball Way

```

Db      838 GACTCAGACAACTCCGCTACTCAGCCGATGACTACAAATPACATGACGGCGCTTT 897
QY      769 AACGACTTGGCGACTGGCTGAGGAGCAAGACGCCAGGCCATCTCCGAAAGGGC 828
Db      898 TCTCCGCTGGCCAGCTTGGCTCGAGCAACAAATGCCAGCTTCTGACAGAAACCGGT 957
QY      829 GCGTCATGAGACCTTGGCTGATGACTGCTTGGCGCCAGAAACAGGCCATTAACGAA 888
Db      958 GGTGGCAAGCTTACGCTCTGATACCAAGACATGTGCCAGCAATCCATATCTCAACGAG 1017
QY      889 AACAGGAGCTCTACATTTGGCTTGTGGGCTGGGCTGCCGGAGCTTTGACAGCTGTAC 948
Db      1018 AACTCAGATGCTATCTTGGCTATGTTGGTGGGCTGCCGAGCATTTGATACAGGTAT 1077
QY      949 ATCTTACTCTGACTCCCTCCGCGAAGCCCGGCACTACACGCA 994
Db      1078 GTCTTGACGGAACACCGACTAGCTGTTACTCATGTGACGACGCA 1123

```

# RESULT 3

US-08-448-873-15  
Sequence 15, Application US/08448873  
Patent No. 5874276

GENERAL INFORMATION:  
APPLICANT: Fowler, Timothy  
APPLICANT: Ward, Michael  
APPLICANT: Clarkson, Kathleen  
APPLICANT: Collier, Katherine A.  
APPLICANT: Larens, Edmund  
TITLE OF INVENTION: No. 5874276e1 Cellulase Enzymes and Systems  
NUMBER OF SEQUENCES: 48  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Genencor International  
STREET: 180 Kimball Way  
CITY: South San Francisco  
STATE: CA  
COUNTRY: USA  
ZIP: 94080

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/448,873  
FILING DATE:  
CLASSIFICATION: 435

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/169,948  
FILING DATE: 17-DEC-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Stone, Christopher L.  
REGISTRATION NUMBER: 35,696  
REFERENCE/DOCKET NUMBER: GC226D14  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 742-7555  
TELEFAX: (415) 742-7217

INFORMATION FOR SEQ ID NO: 15:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1155 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
FEATURE:  
NAME/KEY: CDS  
LOCATION: join(1..56, 231..1155)  
US-08-448-873-15

Query Match 20.9%; Score 275.6; DB 2; Length 1155;  
Best Local Similarity 59.2%; Pred. No. 3.3e-55;

```

Matches 489; Conservative 0; Mismatches 334; Indels 3; Gaps 1;
QY      172 GATGGCGCGGCCAGATGAAAGCATTTGCCGAAGACGAGCGCTCAACGCTTTCGCTA 231
Db      298 GATGGCATGGCCAGATGACGACTTCTGTCACAGAGAGCGGATGATATTTTCGCTTA 357
QY      232 TCCGCTACATGAGCAGTTTGTCTCAACACACGAGTGGAGCGCAAGCTGACGAGCTCAAC 291
Db      358 CCTGTGGATGGAGTAGTACCTCGTCACACAAATTTGGCGGCAATCTTGATTTCCAGAGC 417
QY      292 TGGGGCTCTTACAACAGGTGTCACAGCGCTGTCTCGAGAGCGGCGCTTACTGATAT 351
Db      418 ATTTCCAGATATGATGATGCTTTGTCAGAGGGTGCCTGTCTGCGGCAATCTTCACTGTC 477
QY      352 GACATGACACACTTGGCGGCTTCAACGGGCGCATCATGGCGGAGGAGCGCTGTGGAC 411
Db      478 GACATGCACAAATTAATGCTCATGATGACGAGGATGATGATGATGATGATGATGATGAT 537
QY      412 GACATCTTTGTGACCTTGGCTGGTCCAGATGCAAGACTACTACGAGGACAGCAAGATC 471
Db      538 GCTCAATTCACGAGCCTTTGGTGGCAGTTGGCATCAAGATGACCATCTCAGTGGAGGTG 597
QY      472 ATCTTGGCTGATGAAGACGCGCAGACGCTTGCATGATGATGATGATGATGATGATGAT 531
Db      598 TGGTTCGGCATCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 657
QY      532 CAAGAGGTGTCACGCTGCGATGCGAAGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 591
Db      658 CAAGAGGTGTCACGCTGCGAATGCGAAGCGGCTGCTACGCTGCAATTCATCTTTGCTT 717
QY      592 GGAACCAACTTTGCCAGCGCTGAGAGCTATGCTCACTGCGAGCGGCGGAGCGCTGCGC 651
Db      718 GGAATGATTTGCAATCTGCTGGGCTTTCATATTCGATGCGATGCGATGCGGCGCTGTCT 777
QY      652 AACATTCAGAACCCGAGTGAAGCAGCAGATTTGCTGATCTTGAATGCTCAAGATATCTC 711
Db      778 CAAGTACGAAACCCGAGTGGTCAACAGCAATCTGATTTTTCAGCTGCAAAATCTTG 837
QY      712 GACATCAACAACTCCGGGTGGACGCGGAGTGGACGACCAAGCAACAA--GTGCGAGCGCTTC 768
Db      838 GACTCAGACAACTCCGCTACTCAGCGGAGTACTACAAATPACATGACGGCGCTTT 897
QY      769 AACGACTTGGCGACTGGCTGAGGAGCAAGACGCCAGGCCATCTCCGAAAGGGC 828
Db      898 TCTCCGCTGGCCAGCTTGGCTCGAGCAACAAATGCCAGCTTCTGACAGAAACCGGT 957
QY      829 GCGTCATGAGACCTTGGCTGATGACTGCTTGGCGCCAGAAACAGGCCATTAACGAA 888
Db      958 GGTGGCAAGCTTACGCTCTGATACCAAGACATGTGCCAGCAATCCATATCTCAACGAG 1017
QY      889 AACAGGAGCTCTACATTTGGCTTGTGGGCTGGGCTGCCGAGCTTTGACAGCTGTAC 948
Db      1018 AACTCAGATGCTATCTTGGCTATGTTGGTGGGCTGCCGAGCATTTGATACAGGTAT 1077
QY      949 ATCTTACTCTGACTCCCTCCGCGAAGCCCGGCACTACACGCA 994
Db      1078 GTCTTGACGGAACACCGACTAGCTGTTACTCATGTGACGACGCA 1123

```

## RESULT 4

US-08-382-452D-15  
Sequence 15, Application US/08382452D  
Patent No. 6268196

GENERAL INFORMATION:  
APPLICANT: Fowler, Timothy  
APPLICANT: Clarkson, Kathleen A.  
APPLICANT: Ward, Michael  
APPLICANT: Collier, Katherine D.  
APPLICANT: Larens, Edmund A.  
TITLE OF INVENTION: NOVEL CELLULOSE ENZYMES AND SYSTEMS  
NUMBER OF SEQUENCES: 43  
CORRESPONDENCE ADDRESS:

```

ADDRESSER: Genencor International
STREET: 925 Page Mill Road
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94080
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/382,452D
FILING DATE: February 1, 1995
ATTORNEY/AGENT INFORMATION:
NAME: Christopher L. Stone
REGISTRATION NUMBER: 36,696
REFERENCE/DOCKET NUMBER: GC226-2
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 742-7555
TELEFAX: (415) 742-7217
INFORMATION FOR SEQ ID NO: 15:
SEQUENCE CHARACTERISTICS:
LENGTH: 1155 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: join(1..56, 231..1158)
US-08-382-452D-15

Query Match
Best Local Similarity 20.9%; Score 275.6; DB 4; Length 1155;
Matches 489; Conservative 0; Mismatches 334; Indels 3; Gaps 1;

QY 172 GATGGCGCGCGGAGATGATGATTCGCCGAGAGCGCGGCTTCAAGCTCTTGGCATA 231
D 298 GATGGCGCGCGGAGATGATGATTCGCCGAGAGCGCGGCTTCAAGCTCTTGGCATA 231
QY 232 TCCGCTACATGGCAGTTTGTCTCAACAACAGGTGAGCGGCAAGCTGAGAGCTCAAC 291
D 358 CCTGCGGATGGCAGTACCTGCTCAACAACATTTGGCGGCAATCTGATTTCCAGAGC 417
QY 292 TGGGGCTCTTACACAGGTGTCAAGCCCTCTCGAGAGCGGCGCTTACTGATGAT 351
D 418 ATTTCGAAGTGTGATCAGCTTGTTCAGGGGTGCTCTCTGCGGCGATTCGATCGTC 477
QY 352 GACATGCAACTTGGCCGCTTACAGGGCGGCGATCATCGGCGGAGGGGCTGTCCGAC 411
D 478 GACATCCACATTTATGCTCGATGAGAGCGGTGATCATTTGTCAGGGGCGCTACTAAT 537
QY 412 GACATCTTTTGCACCTCTGGGTCAGATTCGAAAGTACTACGAGCAACAGCAAGATC 471
D 538 GCTCAATTCAGAGCCTTTGGTTCAGATTCGATCAAGTACGATTCGATCGAGGTG 597
QY 472 ATCTTTGGCCGTGATGAGAGCGGCGGAGCTTCGACATTCGATTCGAGCGGCGAGCTGC 531
D 598 TGGTTGGGATCATGATGAGCGGCGGCGGAGCTTCGATTCGATTCGAGCGGCGAGCTGC 657
QY 532 CAAGAAGTGTGATGAGCGGCGGCGGCGGCGGAGCTTCGATTCGATTCGAGCGGCGAGCTGC 591
D 658 CAAGAAGTGTGATGAGCGGCGGCGGCGGCGGAGCTTCGATTCGATTCGAGCGGCGAGCTGC 717
QY 592 GGAACAACCTTTGGCAGCGTTCGAGAGCTATGTCACACTGCGAGCGGAGAGCGGCGGCTGC 651
D 718 GGAAGAATGATGCAATTCGAGAGCGTTCGATTCGATTCGAGCGGCGAGCTGC 777
QY 652 AAGATTGCAAGCGGATGAGAGCGGCGGAGCTTCGATTCGATTCGAGCGGCGAGCTGC 711
D 778 CAAGTCAAGAGCGGATGAGAGCGGCGGAGCTTCGATTCGATTCGAGCGGCGAGCTGC 837

```

```

QY 712 GACATCAACAACCTCGGGTGCAGCGCCGAGTGCACACAGCAAC--GTGACCGCTTTC 768
D 838 GACTAGAACAACTCCGGTACTCAGCGCCGAAATGATTAATTAATGATGACGGCGCTTT 897
QY 769 AACGACTCGGGGACTCGGTGAGCGAACAAGCCGACCATCATCTCCAAACGGGC 828
D 898 TCTCGCTTGGCCACTTGGCTCGAGCAACAATCCGACGATATCTGTGACAAACCGGT 957
QY 829 GCGTCCATGGAACCTTCGTCATGACTGCTTCGCGCCGACGAAAGGCGATTTAGCGAA 888
D 958 GGTGGCAAGCTTCAGTCCGATACAGACATGTCGACGAAATATTCATATTCACACAG 1017
QY 889 AACAGGAGCTTACATTTGCTTTGGCGTGGGGTCCCGGAGCTTTGACACGTCGTAC 948
D 1018 AACTCAGATGTATCTATGCTATGTTGGTGGGTGCGGATCTTGTGATGACGCTAT 1077
QY 949 ATCTGACTGTACCTCCCTCGGCAAGCCCGCACTACACGAGCA .994
D 1078 GTCTGAGGAAACACCGACTAGCTAGTGTATCTGATGAGAGGACA 1123

RESULT 5
US-09-197-649-7
Sequence 7, Application US/09197649
Patent No. 6194550
GENERAL INFORMATION:
APPLICANT: Gold, Larry
APPLICANT: Tuerk, Craig
APPLICANT: Pridnow, David
APPLICANT: Smith, Jonathan D.
TITLE OF INVENTION: Systematic Polypeptide Evolution by Reverse Translation
FILE REFERENCE: ME102/C1-CON
CURRENT APPLICATION NUMBER: US/09/197,649
EARLIER FILING DATE: 1998-11-23
EARLIER APPLICATION NUMBER: 07/829,461
EARLIER FILING DATE: 1992-01-31
EARLIER APPLICATION NUMBER: 07/739,055
EARLIER FILING DATE: 1991-08-01
EARLIER APPLICATION NUMBER: 07/561,968
EARLIER FILING DATE: 1990-08-02
NUMBER OF SEQ ID NOS: 26
SOFTWARE: Patent In Ver. 2.0
SEQ ID NO 7
LENGTH: 390
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: Sequence
OTHER INFORMATION: having a 120 repeat of ACG flanked by fixed
OTHER INFORMATION: fragments having NcoI restriction sites.
US-09-197-649-7

Query Match
Best Local Similarity 4.2%; Score 55.6; DB 4; Length 390;
Matches 178; Conservative 0; Mismatches 204; Indels 0; Gaps 0;

QY 147 GCCCTGCTGAGTACAAAGAGAGATGGCGCGCGAGATGAAAGATTTGCCGAGAGA 206
D 2 GGCATGAGACAGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACG 61
QY 207 CGAGGCGCTCAAGCTTTTCCATATCCGCTACATGCGCATTTTCTCTCAACAACAGGT 266
D 62 CGAGGAGAGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGA 121
QY 267 GAGCGGAGCTGAGCGAGCTCACTGAGGCGCTTACAAAGAGTGTGCAAGCGCTGTCT 326
D 122 CGAGGAGAGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGA 181
QY 327 CGAGGAGGCGGCTTACTGATGATGATGATGATGATGATGATGATGATGATGATGAT 386
D 182 CGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 241
QY 387 CATGCGCCAGGAGGCGGTTCGAGCAGATCTTTGTGACACTTGGGTCCAGATCGCAAA 446

```



```

APPLICATION NUMBER: US 08/218,018
FILING DATE: 23-MAR-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/037,057
FILING DATE: 25-MAR-1993
ATTORNEY/AGENT INFORMATION:
NAME: Pace, Gary M.
REGISTRATION NUMBER: P-40,403
REFERENCE/DOCKET NUMBER: CGC 1695/CIP3/DIV7 - SOLV3
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-541-8582
TELEFAX: 919-541-8689
INFORMATION FOR SEQ ID NO: 42:
SEQUENCE CHARACTERISTICS:
LENGTH: 1241 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: other nucleic acid
DESCRIPTION: /desc = "Synthetic DNA"
FEATURE:
NAME/KEY: CDS
LOCATION: 9..1238
OTHER INFORMATION:
OTHER INFORMATION: sequence encoding VIP2a(a) with the Bacillus secretion signal
OTHER INFORMATION: removed and the eukaryotic secretion signal inserted as
OTHER INFORMATION: contained in pcIB5528"
US-08-471-033-42

```

```

Query Match          4.1%: Score 53.8: DB 1: Length 1241;
Best Local Similarity 46.2%: Pred. No. 0.0009;
Matches 178: Conservative 0; Mismatches 207; Indels 0; Gaps 0;

QY 187 ATGAAGCATTTCCGCGAAGACGAGCGCTCAACGCTTTCCATATCCGCTACATGCGAG 246
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 657 AAGAACGACATCAACGCCGAGGCCACAGCTGGGGCATGAAGAACTACGAGAGATGGGCC 716
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 247 TTTGCTCTCAACAACGAGGTGGAGCGCAAGCTGACGACTCACTGGGGCTCTTACAAC 306
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 717 AAGGACCTGACCGACAGCGACGAGCGGCGCTGGACGGCTACGCCCGCCAGACTACAAAG 776
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 307 AAGTCTGCAACGCGCTGTCTGAGACGCGCCCTACTGACATGATTTGACATGCACACTTT 366
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 777 GAGATCAACAACCTACTGCGAACACGAGGCGGCGGCAAGCGCAAGAACTGGAGCCGAG 836
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 367 GCCCGCTACAACGCGCATCATCGCCAGAGGAGCGGTGCGACGACATCTTTGTGAC 426
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 837 ATCAAGAACATCAGCAGCGCCCTGGGCAAGAGCCCATCCCGAGAACATACCGTGTAC 896
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 427 CTCTGGTCCAGATCGCAAGACTACTACGAGACAACGAAATATCTTTGGCTGATG 486
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 897 CGCTGGTGGCGGATGCCGAGTGTGGCTACCAATCAGCCGATCCCGCCAGCTGTAAG 956
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 487 AAGAGCGCGACGACCTCGACATGAGATCTGGGCGGAGAGCTGCCAAAGAGTGTCACT 546
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 957 GACTTCGAGGAGCATTTCTGTAACACATCAAGAGGCAAGGGCTCATGAGACACGAGC 1016
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 547 GCGATCCGAAAGCGCGGCGCCACT 571
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1017 CTGAGCAGCGAGCGGCTGGCGGCT 1041
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

RESULT 8
US-08-471-044-39
Sequence 39, Application US/084711044
Patent No. 5840868
GENERAL INFORMATION:
APPLICANT: Warren, Gregory W
APPLICANT: Kozielec, Michael
APPLICANT: Mullins, Martha A
APPLICANT: Nye, Gordon J
APPLICANT: Carr, Brian

```

```

APPLICANT: Desai, Nalini M
APPLICANT: Kostichka, N. Kristy
APPLICANT: Duck, Nicholas B
APPLICANT: Estruch, Juan J
TITLE OF INVENTION: No. 5840868el Pesticidal Proteins and Strains
NUMBER OF SEQUENCES: 50
CORRESPONDENCE ADDRESS:
ADDRESS: CIBA-GEIGY Corporation
STREET: 7 Skyline Drive
CITY: Hawthorne
STATE: NY
COUNTRY: USA
ZIP: 10532
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30B
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/471,044
FILING DATE: 06-JUN-1995
CLASSIFICATION: 800
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/463,483
FILING DATE: 05-JUN-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/314,594
FILING DATE: 09-SEP-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/218,018
FILING DATE: 23-MAR-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/037,057
FILING DATE: 25-MAR-1993
ATTORNEY/AGENT INFORMATION:
NAME: Pace, Gary M.
REGISTRATION NUMBER: 40,403
REFERENCE/DOCKET NUMBER: CGC 1695/CIP3/DIV6 - SOLV3
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-541-8582
TELEFAX: 919-541-8689
INFORMATION FOR SEQ ID NO: 39:
SEQUENCE CHARACTERISTICS:
LENGTH: 1241 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: other nucleic acid
DESCRIPTION: /desc = "Synthetic DNA"
FEATURE:
NAME/KEY: CDS
LOCATION: 9..1238
OTHER INFORMATION:
OTHER INFORMATION: sequence encoding VIP2a(a) with the Bacillus secretion sig
OTHER INFORMATION: removed as contained in pcIB5527"
US-08-471-044-39

```

```

Query Match          4.1%: Score 53.8: DB 2: Length 1241;
Best Local Similarity 46.2%: Pred. No. 0.0009;
Matches 178: Conservative 0; Mismatches 207; Indels 0; Gaps 0;

QY 187 ATGAAGCATTTCCGCGAAGACGAGCGCTCAACGCTTTCCATATCCGCTACATGCGAG 246
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 657 AAGAACGACATCAACGCCGAGGCCACAGCTGGGGCATGAAGAACTACGAGAGATGGGCC 716
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 247 TTTGCTCTCAACAACGAGGTGGAGCGCAAGCTGACGACTCACTGGGGCTCTTACAAC 306
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 717 AAGGACCTGACCGACAGCGGCGGCGGCGCTGGACGGCTACGCCCGCCAGAGACTAAG 776
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 307 AAGTCTGCAACGCGCTGTCTGAGACGCGGCGCTACTGACATGATTTGACATGCACACTT 366
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 777 GAGATCAACAACCTACTGCGAACACGAGGCGGCGGCAAGCGCAAGAACTGGAGCGCCAG 836
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

OY 367 GCCCGCTACAAGCGCGCATCATCGCCAGAGGCGGTGTGCGACGATCTTTGCGAC 426
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 837 ATCAAAACATCATGACGACCGCCCTGGGCGACAGACCCATCCCGAAGACATCACCCTGTAC 896
OY 427 CTCTGGTCCAGATCGCAAGTACTAGAGACAGACAGATCTTTGGCTGTAG 486
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 897 CGCTGGTGGCATGCGCCGAGTTCGGCTACCAAGATCAGCGACCCCTGGCCACCTGTAG 956
OY 487 AACGAGCCCGACGACCTCGACATTTGAGATCTGGGCGCAGACGTCGCAAAAGTCTGACT 546
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 957 GACTTCGAGAGAGAGTTCCTGAAACACCATCAAGAGAGACAAAGGCTTACATGAGACACG 1016
OY 547 GCGATCCGAAGGCGCGCCGACCT 571
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1017 CTGAGCAGCGAGCGCTGGCGGCT 1041

```

## RESULT 9

```

US-08-471-044-42
; Sequence 42, Application US/08471044
; Patent No. 5840868
; GENERAL INFORMATION:
; APPLICANT: Warren, Gregory W
; APPLICANT: Koziel, Michael G
; APPLICANT: Mullins, Martha A
; APPLICANT: Nye, Gordon J
; APPLICANT: Carr, Brian
; APPLICANT: Desai, Nalini M
; APPLICANT: Kostichka, N. Kristy
; APPLICANT: Duck, Nicholas B
; APPLICANT: Estruch, Juan J
; TITLE OF INVENTION: No. 5840868el Pesticidal Proteins and Strains
; NUMBER OF SEQUENCES: 30
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CIBA-GEIGY Corporation
; STREET: 7 Skyline Drive
; CITY: Hawthorne
; STATE: NY
; COUNTRY: USA
; ZIP: 10532
; COMPUTER READABLE FORM:
; MEDIUM TYPE: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.308
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/471,044
; FILING DATE: 06-JUN-1995
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/463,483
; FILING DATE: 05-JUN-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/314,594
; FILING DATE: 09-SEP-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/218,018
; FILING DATE: 23-MAR-1994
; APPLICATION DATA:
; APPLICATION NUMBER: US 08/037,057
; FILING DATE: 25-MAR-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Pace, Gary M.
; REGISTRATION NUMBER: 40,403
; REFERENCE/DOCKET NUMBER: CGC 1695/CIP3/DIV6
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-541-8582
; TELEFAX: 919-541-8689
; INFORMATION FOR SEQ ID NO: 42:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1241 base pairs
; TYPE: nucleic acid

```

SQLV3

```

; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: other nucleic acid
; DESCRIPTION: /desc = "Synthetic DNA"
; HYPOTHETICAL: NO
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 9..1238
; OTHER INFORMATION: /note="Maize optimized DNA
; OTHER INFORMATION: sequence encoding VIP2(a) with the Bacillus secretion sig
; OTHER INFORMATION: removed and the eukaryotic secretion signal inserted as
; OTHER INFORMATION: contained in pCIB5528"
US-08-471-044-42

```

```

Query Match 4.1%; Score 53.8; DB 2; Length 1241;
Best Local Similarity 46.2%; Pred. No. 0.0009;
Matches 178; Conservative 0; Mismatches 207; Indels 0; Gaps 0;

```

```

OY 187 ATGAGCATTTCCCGAAGACGAGCGGCTCAACGCTTTTGCATATCCGCTACATGCGAC 246
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 657 AAGAGCAGCATCAACGCCGAGGCCACACAGCTGGGCGATGAAGACTACGAGAGTGGCC 716
OY 247 TTTGTCTCAACAACACAGCGGTGACGGCAGCTGGAGAGCTCAACTGGGCTCTACAC 306
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 717 AAGAGCTGACCGACAGCGCGAGCGGAGGCGCTGAGCGGCTTACGCGCCGACGACTAC 776
OY 307 AAGTGTCAAGCGCTGTCTCGAGAGCGGCGCTACTGCTATGATTGACATGCACACTTT 366
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 777 GAGATCAACAACATCACTGCTGCCCAACAGGCGGAGCGGCAAGAGAGCTGGAGCGCC 836
OY 367 GCCCGCTACAAGCGCGGCGCATCATCGGCGAGGAGCGGTGTGCGACGATCTTTGCGAC 426
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 837 ATCAAGACATCATGACGACCGCCCTGGGCGACAGACCCATCCCGAAGACATCACCCTGTAC 896
OY 427 CTCTGGTCCAGATCGCAAGTACTAGAGACAGACAGATCATCTTTGGCTGTAG 486
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 897 CGCTGGTGGCATGCGCCGAGTTCGGCTACCAAGATCAGCAACCCCTGGCCACCTGTAG 956
OY 487 AACGAGCCCGACGACCTCGACATTTGAGATCTGGGCGCAGACGTCGCAAAAGTCTGACT 546
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 957 GACTTCGAGAGAGAGTTCCTGAAACACCATCAAGAGAGACAAAGGCTTACATGAGACACG 1016
OY 547 GCGATCCGAAGGCGCGCCGACCT 571
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1017 CTGAGCAGCGAGCGCTGGCGGCT 1041

```

## RESULT 10

```

US-08-463-483A-39
; Sequence 39, Application US/08463483A
; Patent No. 5849870
; GENERAL INFORMATION:
; APPLICANT: Warren, Gregory W
; APPLICANT: Koziel, Michael G
; APPLICANT: Mullins, Martha A
; APPLICANT: Nye, Gordon J
; APPLICANT: Carr, Brian
; APPLICANT: Desai, Nalini M
; APPLICANT: Kostichka, N. Kristy
; APPLICANT: Duck, Nicholas B
; APPLICANT: Estruch, Juan J
; TITLE OF INVENTION: No. 5849870el Pesticidal proteins and Strains
; NUMBER OF SEQUENCES: 50
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CIBA-GEIGY Corporation
; STREET: 7 Skyline Drive
; CITY: Hawthorne
; STATE: NY
; COUNTRY: USA
; ZIP: 10532
; COMPUTER READABLE FORM:
; MEDIUM TYPE: IBM PC compatible

```

```

OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.308
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/463,483A
FILING DATE:
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/314,594
FILING DATE: 09-SEP-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/218,018
FILING DATE: 23-MAR-1994
APPLICATION NUMBER: US 08/037,057
FILING DATE: 25-MAR-1993
ATTORNEY/AGENT INFORMATION:
NAME: Spruill, W. Murray
REGISTRATION NUMBER: 32,943
REFERENCE/DOCKET NUMBER: CGC 1695/CIP3
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-541-8615
TELEFAX: 919-541-8689
INFORMATION FOR SEQ ID NO: 39:
SEQUENCE CHARACTERISTICS:
LENGTH: 1241 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: other nucleic acid
DESCRIPTION: /desc = "synthetic DNA"
HYPOTHETICAL: NO
FEATURE:
NAME/KEY: CDS
LOCATION: 9..1238
OTHER INFORMATION: /note= "Maize optimized DNA
OTHER INFORMATION: sequence encoding VIP2A(a) with the Bacillus secretion signal
US-08-463-483A-39

```

```

Query Match 4.1%; Score 53.8; DB 2; Length 1241;
Best Local Similarity 46.2%; Pred. No. 0.0009;
Matches 178; Conservative 0; Mismatches 207; Indels 0; Gaps 0;

OY 187 ATGAAGCATTTCCGCGAAGAGAGAGAGGCGCTCAACGCTTTCCGATATCCGCTACATGCGAG 246
Db 657 AAGAAGCATATCAAGCGCCGAGGCGCCACAGCTGGGGCATGAAGAAGTACGAGAGTGGGCC 716
OY 247 TTTGCTCTCAACAACACGCGTGGAGCGCAAGCTGAGAGCTCACTGAGGCTCTCAACAAC 306
Db 717 AAGGACCTGACCGAAGCAGCAGCGAGGCGCTGAGCGCTACGCCCGCAGAGACTACAAAG 776
OY 307 AAGTCTGCAACGCTGCTCTGAGAGAGGCGCTTACTGATGATGACATGCACAACCTT 366
Db 777 GAGATCAACAACCTACTGCGCAACACGAGGCGGCGCAAGAGAACTGAGAGCGCCAG 836
OY 367 GCCGCTACAGCGGCGGATCATGCGCAGGAGAGGCGCTGTGGAGAGATCTTTGTCCAGC 426
Db 837 ATCAAGACATACAGCGAGCGCTGGGCAAGAAGCCCATCCCGAAGACATACCGCTGAC 896
OY 427 CTCTGGTTCAGATCGCAAGATAGTACGAGAGCAACGACAGATCATTTGGCCTGATG 486
Db 897 CGCTGGTGGCGGATGCCCGATTTGGGCTACCAAGATCAAGCGAGCCCTGCGAGCTGAAG 956
OY 487 AAGCAGCCGACGACCTGACATTTGAGATCTGGGCGGCGAGAGCTGCCAAGAAGGTGTCAC 546
Db 957 GACTTCGAGAGGAGCTTCTCAACACCATCAAGAGGAGCAAGGGCTACATGAGACCGAGC 1016
OY 547 GCGATCCGAAGAGCGCGCGCCACCT 571
Db 1017 CTGAGAGCGAGCGCTGGCGCGCT 1041

```

```

US-08-463-483A-42
Sequence 42, Application US/08463483A
Patent No. 5849870
GENERAL INFORMATION:
APPLICANT: Warren, Gregory W
APPLICANT: Koziele, Michael G
APPLICANT: Mullins, Martha A
APPLICANT: Nye, Gordon J
APPLICANT: Carr, Brian
APPLICANT: Desai, Nalin M
APPLICANT: Kostichka, N. Kristy
APPLICANT: Duck, Nicholas B
APPLICANT: Estruch, Juan J
TITLE OF INVENTION: No. 5849870e1 Pesticidal Proteins and Strains
NUMBER OF SEQUENCES: 50
CORRESPONDENCE ADDRESS:
ADDRESS: CIBA-GEIGY Corporation
STREET: 7 Skyline Drive
CITY: Hawthorne
STATE: NY
COUNTRY: USA
ZIP: 10532
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.308
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/463,483A
FILING DATE:
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/314,594
FILING DATE: 09-SEP-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/218,018
FILING DATE: 23-MAR-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/037,057
FILING DATE: 25-MAR-1993
ATTORNEY/AGENT INFORMATION:
NAME: Spruill, W. Murray
REGISTRATION NUMBER: 32,943
REFERENCE/DOCKET NUMBER: CGC 1695/CIP3
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-541-8615
TELEFAX: 919-541-8689
INFORMATION FOR SEQ ID NO: 42:
SEQUENCE CHARACTERISTICS:
LENGTH: 1241 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: other nucleic acid
DESCRIPTION: /desc = "synthetic DNA"
HYPOTHETICAL: NO
FEATURE:
NAME/KEY: CDS
LOCATION: 9..1238
OTHER INFORMATION: /note= "Maize optimized DNA
OTHER INFORMATION: sequence encoding VIP2A(a) with the Bacillus secretion sig
OTHER INFORMATION: removed and the eukaryotic secretion signal inserted as
US-08-463-483A-42

```

```

Query Match 4.1%; Score 53.8; DB 2; Length 1241;
Best Local Similarity 46.2%; Pred. No. 0.0009;
Matches 178; Conservative 0; Mismatches 207; Indels 0; Gaps 0;

OY 187 ATGAAGCATTTCCGCGAAGAGAGAGGCGCTCAACGCTTTCCGATATCCGCTACATGCGAG 246
Db 657 AAGAAGCATATCAAGCGCCGAGGCGCCACAGCTGGGGCATGAAGAAGTACGAGAGTGGGCC 716

```

;; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: No. 5866326artis Corporation



```

; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/463,483
;

```

Search completed: May 15, 2003, 13:49:20  
Job time : 64.7022 secs



GenCore version 5.1.4\_p5\_4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 15, 2003, 13:49:29 ; Search time 159.649 Seconds  
(without alignments)  
10641.144 Million cell updates/sec

Title: US-10-028-245-4  
1317

Sequence: 1 atgcgcgaacctccctct.....acatgctgtagcgctttga 1317

Scoring table: IDENTITY\_NUC

Searched: . 810007 seqs, 644969091 residues

Total number of hits satisfying chosen parameters: 1620014

```
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
```

```
Post-processing: Minimum Match 0%
                  Maximum Match 100%
```

Listing first 45 summaries

Database : Published\_Applications\_NA:\*.

- 1: /cgn2\_6/prodata/2/pubnpa/US07\_PUBCOMB.seq: \*
- 2: /cgn2\_6/prodata/2/pubnpa/PCT\_NEW\_PUB.seq: \*
- 3: /cgn2\_6/prodata/2/pubnpa/US06\_NEW\_PUB.seq: \*
- 4: /cgn2\_6/prodata/2/pubnpa/US06\_PUBCOMB.seq: \*
- 5: /cgn2\_6/prodata/2/pubnpa/US07\_NEW\_PUB.seq: \*
- 6: /cgn2\_6/prodata/2/pubnpa/PCTUS\_PUBCOMB.seq: \*
- 7: /cgn2\_6/prodata/2/pubnpa/US08\_NEW\_PUB.seq: \*
- 8: /cgn2\_6/prodata/2/pubnpa/US08\_PUBCOMB.seq: \*
- 9: /cgn2\_6/prodata/2/pubnpa/US09\_NEW\_PUB.seq: \*
- 10: /cgn2\_6/prodata/2/pubnpa/US09\_PUBCOMB.seq: \*
- 11: /cgn2\_6/prodata/2/pubnpa/US10\_NEW\_PUB.seq: \*
- 12: /cgn2\_6/prodata/2/pubnpa/US10\_PUBCOMB.seq: \*
- 13: /cgn2\_6/prodata/2/pubnpa/US60\_NEW\_PUB.seq: \*
- 14: /cgn2\_6/prodata/2/pubnpa/US60\_PUBCOMB.seq: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

## SUMMARIES

Table 1: Performance Metrics						Description
Result No.	Score	Query Match	Length	DB	ID	
1	275.6	20.9	1155	9	US-09-916-496A-15	Sequence 15, Appl
2	132	10.0	2957	9	US-09-927-827-20	Sequence 20, Appl
3	75.8	5.8	3002	9	US-09-927-827-25	Sequence 25, Appl
4	65.8	5.0	2040	9	US-09-927-827-26	Sequence 26, Appl
5	55.6	4.2	390	10	US-09-790-399-7	Sequence 7, Appl
6	51.8	3.9	185	10	US-09-887-576-784	Sequence 784, Appl
7	47	3.6	1294	10	US-09-748-033-2	Sequence 2, Appl
8	46.8	3.6	520	9	US-10-184-634-332	Sequence 332, App
C	46.8	3.6	520	9	US-10-184-634-332	Sequence 332, App
C	46	3.5	1347	10	US-09-350-756-6	Sequence 332, App
10	45.8	3.5	905	10	US-09-748-033-7	Sequence 6, Appl
11	45.8	3.5	905	10	US-09-748-033-5	Sequence 7, Appl
12	45.6	3.5	905	10	US-09-748-033-5	Sequence 6, Appl
13	45.4	3.4	368	9	US-09-988-462-2	Sequence 5, Appl
14	45.2	3.4	1371	10	US-09-910-186A-9	Sequence 2, Appl
15	45.2	3.4	1371	10	US-09-350-756-3	Sequence 9, Appl
16	45.2	3.4	1734	9	US-09-738-626-1062	Sequence 3, Appl
17	44.4	3.4	1098	9	US-09-815-242-7874	Sequence 1062, App
18	44.4	3.4	1368	9	US-09-910-186A-17	Sequence 7874, App
19	44.4	3.4	1593	10	US-09-997-664-102	Sequence 17, Appl
						Sequence 102, Appl

20	44.4	3.4	6.91	10	US-09-997-664-1	Sequence 1, Appl1
21	43.8	3.3	11.07	10	US-09-748-033-6	Sequence 6, Appl1
22	43.8	3.3	27.12	10	US-09-748-033-4	Sequence 4, Appl1
23	43.8	3.3	34.68	9	US-09-988-462-4	Sequence 4, Appl1
24	43	3.3	64.42	9	US-09-950-335A-11	Sequence 11, Appl
25	42.2	3.3	35.96	9	US-09-988-462-14	Sequence 14, Appl
26	42	3.2	33.60	10	US-09-750-399-7	Sequence 7, Appl1
27	41.8	3.2	7.41	10	US-09-815-242-7843	Sequence 7843, App
28	41.8	3.2	13.62	9	US-09-970-966-208	Sequence 208, App
29	41.8	3.2	13.62	9	US-10-097-340-118	Sequence 118, App
30	41.8	3.2	13.62	10	US-09-825-234-208	Sequence 208, App
31	41.8	3.2	17.85	10	US-09-864-761-18040	Sequence 18040, App
32	41.8	3.2	24.28	9	US-09-882-849A-6	Sequence 6, Appl1
33	41.8	3.2	36.24	9	US-09-988-462-6	Sequence 6, Appl1
34	41.6	3.2	12.84	9	US-10-086-738A-2	Sequence 2, Appl
35	41.6	3.2	12.84	9	US-10-086-738A-3	Sequence 3, Appl
36	41.4	3.1	8.67	10	US-09-216-393-440	Sequence 340, Appl
37	41.4	3.1	8.67	10	US-09-216-393-440	Sequence 342, Appl
38	41.4	3.1	13.97	10	US-09-216-393-443	Sequence 343, App
39	41.4	3.1	13.97	10	US-09-216-393-445	Sequence 345, App
40	41.4	3.1	16.56	12	US-10-033-109-13	Sequence 13, Appl
41	41	3.1	4.77	9	US-09-938-842A-2589	Sequence 289, App
42	41	3.1	6.81	10	US-09-770-149-341	Sequence 341, App
43	41	3.1	14.64	9	US-10-166-087-15	Sequence 15, Appl
44	41	3.1	32.59	9	US-10-166-087-1	Sequence 1, Appl
45	40.8	3.1	7.05	9	US-10-213-990-68	Sequence 68, Appl

## ALIGNMENTS

## RESULT 1

US-09-916-494A-15  
; Sequence 15, Application US/09916494A

; GENERAL INFORMATION:

APPLICANT: Clarkson, Kathleen A.

APPLICANT: Collier, Katherine D.

10; TITLE OF INVENTION: Method and c

1. TITLE OF INVENTION: Enzyme compositions

CURRENT APPLICATION NUMBER: US/09/916,494A

PRIOR APPLICATION NUMBER: US 08/382,452

; PRIOR APPLICATION NUMBER: US 08/169,948

NUMBER OF SEQ ID NOS: 43

; SEQ ID NO 19

TYPE: DNA

**FEATURE:**

; LOCATION: (1)...(56)

LOCATION: (231)...(1155)

100

Best Local Similarity 59

0  
1  
2  
3  
4  
5  
6  
7  
8  
9  
A  
B  
C  
D  
E  
F  
G  
H  
I  
J  
K  
L  
M  
N  
O  
P  
Q  
R  
S  
T  
U  
V  
W  
X  
Y  
Z

[illegible][illegible]

```

Db 358 CCTGCGGATGGAGTACCTGTCACAAATTTGGGCGGCGCATCTTGATTCACGAGC 417
QY 292 TGGGGCTCTACACAAAGCTGTCAAGGCTGTCTCGAGAGGGGCGCTTACGTAT 351
Db 418 ATTCCAAAGTATGATCAGCTTGTTCAGGGGTGCTGTCTGGGGCGCATCTGCTGCT 477
QY 352 GACATGCAACTTTGGCGGCTACAAAGCGGCGCATCATCGGCGAGGAGGGGTGCGAC 411
Db 478 GACATCCCAATTTATGCTGATGGAAGCGGTGGATCATTTGCGAGGGGCGCTACTAAT 537
QY 412 GACATCTTTGTCGACCTTGGGTCCAGATCCAAAGTACTTACGAGACACAGCAAGTTC 471
Db 538 GCTCAATTCAGAGCCTTTGGTCCGAGATGGCATCAAGTACGCAATCTCATCTCGAGGTG 597
QY 472 ATCTTTGGGCTGTATGAACGAGCGCGGACCTCGACATTTGATGTGGCGGCGAGCGTG 531
Db 598 TGTGTCGGCATCATGATGATGAGCCCGACGACGTGAACATCAACACCTGGGCTGCCAGGTG 657
QY 532 CAAAAGTCTGCTACGTGCGATCGAAAGCGCGCGCACCTGCGAGATGATCTCTGCGCC 591
Db 658 CAAGAGTGTATACCGCAATCCGAAAGCGTGTGCTAGCTGCAATTCATCTCTTGTGCT 717
QY 592 GGAACCAACTTGGCAGGTGAGACGATGTGTCTACCTGCGAGCGCGGAAGCCCTCGGC 651
Db 718 GGAATGATTTGGCAATCTGCTGGGCTTTCAATATCCGATGCGAGTGCAGCGCCCTGTCT 777
QY 652 AAGATATGCAACCGGATGAGAAACCGATTTGCTGCTACTTTGATGTGCACAGATCTC 711
Db 778 CAAGTCAAGAACCGGATGGGTCAACAAAGCAATCTGATTTTGTGAGCTGACAAATACTG 837
QY 712 GACATCAACAACCTCGGGTCCAGCGCGAGTGCACACAGCAAC--GTGAGCGCTTTC 768
Db 838 GACTCAGACAACTCCGCTACTACGCGCGAAATGTAACAAATTAACATTTGAGCGGCTTT 897
QY 769 AAGCACTCGGAGCTGGCTGAGGAGAAAGCGCGGCGCATCATCTCGAAAGCGGC 828
Db 898 TCTCGGCTTGGCACTTGGCTCGAGACAAATGCGCGCTATCTTACAGAAACCGGT 957
QY 829 GCGTCATGGAACCTTCTGCTGATGACTGCTTTCGCGCGCGCAACAAAGCGCATTTAGCGAA 888
Db 958 GGTGGCAACGTTCACTGCTGATGATGACATGATGCGCAAAATTCATATCTCAACAG 1017
QY 889 AAGAGCGAGCTCTACATTTGGCTTGTGGGTGGGGTGGCGGCGAGCTTTGACAGCTGTAC 948
Db 1018 AACTCAGATGCTATCTTGGCTATGTTGGTGGGGTGGCGGATCATTTGATAGACGTAT 1077
QY 949 ATCTTGACTGTGACTCCCTCGGCAAGCGCGCAACTACACCGACA 994
Db 1078 GTCTGACGGAACACGAGTACGAGTGTACTCATGAGCGGACA 1123

```

## RESULT 2

```

US-09-927-827-20/c
: Sequence 20, Application US/09927827
: Publication No. US20030036176A1
: GENERAL INFORMATION:
: APPLICANT: Bower, Stanley G.
: APPLICANT: Ramseler, Thomas M.
: TITLE OF INVENTION: Directed Genetic Engineering of Xanthomonas campestris
: FILE REFERENCE: 38-10(15824)B
: CURRENT APPLICATION NUMBER: US/09/927,827
: PRIOR FILING DATE: 2001-08-10
: PRIOR APPLICATION NUMBER: US 60/279,493
: NUMBER OF SEQ ID NOS: 69
: SEQ ID NO 20
: LENGTH: 2957
: TYPE: DNA
: ORGANISM: Xanthomonas campestris
: FEATURE:
: NAME/KEY: CDS
: LOCATION: (1001)..(1957)

```

## US-09-927-827-20

Query Match 10.0%; Score 132; DB 9; Length 2957;

Best Local Similarity 51.5%; Pred. No. 2,2e-30; Matches 384; Conservative 0; Mismatches 350; Indels 12; Gaps 3;

```

QY 252 CCTCAACAACAGGTGAGCGGCAAGCTGAGAGCTCAATGAGGCTCTTACAACAAGT 311
Db 1793 CCTGAGCGCAACCTCATATGCGAGCTGATGCGGCGCACTGCTGATCAAAAATTC 1734
QY 312 CGTCAAGCCTGTCTCGAGAGGGGCGCTACTGATATTTACATGCAACTTTGCCCG 371
Db 1733 GGTGGCGCGCGCAAGGCAACAGCACTTCTGATCTTGGACCAACTACGCGAA 1674
QY 372 CTACAAAGCGGCGATCATCGCGAGGAGGCTGTGCGACACATCTTTGCGACTCTG 431
Db 1673 GTACACAGGAAGCGCATCGGACCGACGAGGTGCGGCTCGACACTGGCGCATCTGTG 1614
QY 432 GGTCCAGATGCGAAAGTACTACAGAGACACAGCAAGATCATCTTTGGCTGATGACGA 491
Db 1613 GCGCGCGCTGGCGCTGAGTTCAAAAGACAGCAAGAGTGTATCTTGGGCTGATGAAACGA 1554
QY 492 GCGCGAGACCTCGACATTTGAGATCTGCGCGCAGACAGCTGCCAAAAGTCTGACGCT 551
Db 1553 GCCCAATGCACTCTCTCCACCGACTGGCGCGCGCGCAACGCGGATCAACACCAT 1494
QY 552 CCGAAAGCGCGCGCGCACCTGCGAGATGATCTCTGCGCGCGCAACCAACTTTGCGAGCT 611
Db 1493 CCGCAAGCGCGCGCGGAGAAC--AATTGATTTCTGTGCTGCGAGGCTCTACACCGGCG 1437
QY 612 CGAGACGATGTGTCTACTGCGAGGCGGAGAGCCCTGCGCAAG--ATTACGAACCGGA 668
Db 1436 GCACAGCTGGCGGACAGCACAGCTGAGGCTCTCAATGCGCAAGGCGCGCTGAGATCTCAA 1377
QY 669 TGAAGACGCGATTTGCTGCTGATGATGTCACAAAGTATCTGCAATCAACACTCGG 728
Db 1376 GAGCCCGCGCAACATTTGCGCTTTCGAGCGCATCATATCTGAGCGGCTACAGCGG 1317
QY 729 GTGCGACCGCGAGTGCACACAGACAGACGTCGAGCGCTTCAA-----CGACTTGGCGGA 782
Db 1316 CACCAAGCGCGAATATGCGCTTTCGAGCGCATCATATCTGAGCGGCTACAGCGG 1257
QY 783 CTGCTGAGCGAGAACAAAGCGCGAGCATATCTTCGAAAGGGGCGGTCTCATGGAAC 842
Db 1256 CTGCTGCGCGAGAACAAAGCGCTTCTGCGGCGAGTTTCGCGACGCGCAACATGCT 1197
QY 843 TTGCGATGACTGCTGCTTCTGCGCGCAGAACAGGCGCTTAGCGAAACAGCAGCTTA 902
Db 1196 GGTCTGACAGACGCGCTGAAAGGCGATGCTGAGCTACATGGAAGAAACAGCGAGCTGTG 1137
QY 903 CATTGGCTTTGTGGGCTGGGGTGGCGGAGCTTTGACAGCTGTACATCTTGACTGTGAC 962
Db 1136 GCTGGGCTGTGACCTGTGTGGGCGCGGCGGTGTGGAAGCGGAGTACCGCTTCAATGT 1077
QY 963 TCCCTCGGCAAGCGCGGCACTTACA 988
Db 1076 GCAGCGCGGCAAGAGCGGCGACGAGGA 1051

```

## RESULT 3

```

US-09-927-827-25
: Sequence 25, Application US/09927827
: Publication No. US20030036176A1
: GENERAL INFORMATION:
: APPLICANT: Bower, Stanley G.
: APPLICANT: Ramseler, Thomas M.
: TITLE OF INVENTION: Directed Genetic Engineering of Xanthomonas campestris
: FILE REFERENCE: 38-10(15824)B
: CURRENT APPLICATION NUMBER: US/09/927,827
: PRIOR FILING DATE: 2001-08-10
: PRIOR APPLICATION NUMBER: US 60/279,493
: NUMBER OF SEQ ID NOS: 69

```

```

: SEQ ID NO 25
: LENGTH: 3002
: TYPE: DNA
: ORGANISM: Xanthomonas campestris
: FEATURE:
: NAME/KEY: CDS
: LOCATION: (1001)..(2002)
US-09-927-827-25

```

Query Match	5.8%	Score 75.8	DB 93	Length 3002
Best Local Similarity	46.7%	Pred. No. 4.8e-13		
Matches 279	Conservative 0	Mismatches 312	Indels 6	Gaps 1

QY 340 TACGCGATGATTACATATGCACCACTTTGGCCGCTACAGCGGGCATATCGGCCAGGGA 399

Db 1298 TACTGATTATTCATGTGTGCACCACTACGCCAAGTACTACGGCCAGAAAGTTGGCAGTAA 1357

QY 400 GCGGTGTGAGACACATCTTTGTGCACCTTGTGGTCCAGATCGCAAGTACTACGAGAC 459

Db 1358 CGGGTGGCGGTGGCGACAGTTCAACCGATCTGTGGCGCCGCTGTGGCCATGGCGTTCAAGAGC 1417

QY 460 AAGCAAGAATCATCTTTGGCTGTATGAACGAGCGGCACGACTTGACATTGAGATCTGG 519

Db 1418 GACAACCCCTGATTTTTCGGGGTGATGAACGAGCCGTACGACATCTCGCCAGAGAGTTGG 1477

QY 520 GCGCAGACGTGCCAAAAGTCGCTACTGGGATGCCGAAGACGGGGGCCACCTCGGAGATG 579

Db 1478 GCGCGCGCGCGCGACGGCGTCAATCGACAGCATTTCTGGCCACCGGGTCCACCAACTGTGATT 1537

QY 580 ATCTCTCTCGCCCGGAACCAACTTTGGCAGCGTGCAGACGTATGTGTCCACTGGCAGCCG 639

Db 1538 CTGGTGGCGGGTGCATTGTGTGTGGGGCGCACAGCTGTATTTCACCGTGGCCGGGAG 1597

QY 640 GAAGCCCTGGCAAGATTACGAACCCGGATGGAAGCACCGCATTTGCTTACTTTGATGTC 699

Db 1598 TCCAAATGCAATGGCGGCTTGGCAATCCCGACCCATTCGAATGCGTATTCGATCGAAGTC 1657

QY 700 CACAGATATCGCAATCAACCAACTCGGGGTGGCAGCGGCAATGCAACAGACAAGTCTC 759

Db 1658 CATCAATATCTGGACACCGCATCCAGCGACCAAGTCCGGTTGTGTGAGCCGCAAGTAT 1717

QY 760 GAGC-----CTTCAACGACTTGGGAGATGGCTGAGCGCAACAAGCGCCAGGCAATC 813

Db 1718 GGTGGGAGGCGCTTGGGCACATTTACCGGGGTGGCTCGCGGCCAAGCAAGCGCGGTTTT 1777

QY 814 ATCTCGAAGAGGCGCGCTCATGGAACCTTGTGTGATGACTTGCCTTCTGCGCCCAAGAC 873

Db 1778 CTTTGGCAATTCGGCACCGGCACCAATTAACGAACCTGACATCGCGGCTGGACGGCATGCTC 1837

QY 874 AAGGCCATTGCGAAAACACCGACGCTCATATTGGCTTTGTGGCTGGGGTGTCCGCG 930

Db 1838 GTTTATCTCGAACAACCAACCGATGTGTGATTCGATTGGAGCTTCTGTGGCGGCGAGCG 1894

```

? RESULT 4
? US-09-927-827-26
? :
? :
? : Sequence 26. Application US/09927827
? : Publication NO. US20030036176A1
? :
? : GENERAL INFORMATION:
? :
? : APPLICANT: Bower, Stanley G.
? : APPLICANT: Ramsesler, Thomas M.
? : TITLE OF INVENTION: Directed Genetic Engineering of Xanthomonas campestris
? : FILE REFERENCE: 38-10(15824)B
? :
? : CURRENT APPLICATION NUMBER: US/09/927,827
? :
? : CURRENT FILING DATE: 2001-08-10
? :
? : PRIORITY APPLICATION NUMBER: US 60/279,493
? :
? : PRIORITY FILING DATE: 2001-03-28
? :
? : NUMBER OF SEQ ID NOS: 69
? :
? : SEQ ID NO 26
? :
? : LENGTH: 2040
? :
? : TYPE: DNA
? :
? : ORGANISM: Xanthomonas campestris
? :
? : FEATURE:
? :

```

```

; NAME/KEY: CDS
; LOCATION: (1001)..(2008)
US-09-927-827-26

```

Query Match	5.0%;	Score 65.8;	DB 9;	Length 2040;
Best Local Similarity	53.3%;	Pred. No. 5e-10;		
Matches 139;	Conservative 0;	Mismatches 122;	Indels 0;	Gaps 0;

QY	332	CGCGCGCTTCATGATGATGACATGCACAACCTTTGCCGGTACCAAGCGGCGCATCTTCG	391
Db	1293	CGGGCATGTATCTGTGTCTGTGATATCCACAACTACAGCAAGTACTTCGGTTTAAATAATG	1353
QY	392	GCCAGGAGGCGGTCTGGAGCAGCATCTTTGTGCACCTGTGGGTCCAGATCGCAAGTACT	451
Db	1353	GTTGGGCGGAAGTGCCTGCTGCCACCTTTGGCATCTGTGGCGGCGCTGGCGGGATCT	1412
QY	452	ACGAGCAACAGCAACAGATATCTTTGGCGTGTGAAGCGCGGACGACGCTCGACATTTG	511
Db	1413	TCAACAGCGTAAACGCCGTGATCTTTGGGCTGTATGTAAGCCCAACACATCTTGCCA	1472
QY	512	AGATCTTGGCGGAGAGCGTGGCCAAAAGGTGCTACTGATTCGCAAAAGCGCGGCGCACT	571
Db	1473	GCGATGTGGCGGTGTCAGCCACGACGAGCATGATCGATCCGTGCACCGCGGCCACACA	1532
QY	572	CGCAGTATCTCTCTCCCGG	592
Db	1533	ATCTGATCTGTACCGGGCG	1553

RESULT 5  
US-09-790-399-7  
; Sequence 7, Application US/09790399  
; Patent No. US20020038000A1  
; GENERAL INFORMATION:

```

APPLICANT: Pribnow, David
APPLICANT: Smith, Jonathan D.
TITLE OF INVENTION: Systematic Polypeptide Evolution by Reverse Translation
FILE REFERENCE: NEX02/C1-CON2
CURRENT APPLICATION NUMBER: US/09/790,399
CURRENT FILING DATE: 2001-02-22
PRIORITY APPLICATION NUMBER: 09/197,649
PRIORITY FILING DATE: 1998-11-23
PRIORITY APPLICATION NUMBER: 07/829,461
PRIORITY FILING DATE: 1992-01-31
PRIORITY APPLICATION NUMBER: 07/739,055
PRIORITY FILING DATE: 1991-08-01
PRIORITY APPLICATION NUMBER: 07/561,968
PRIORITY FILING DATE: 1990-08-02
NUMBER OF SEQ ID NOS: 26
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 7
LENGTH: 390
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: Sequence
OTHER INFORMATION: having a 120 repeat of AGC flanked by fixed
OTHER INFORMATION: fragments having NcoI restriction sites.
US-09-790-399-7

```

[illegible]





RESULT 13  
US-09-988-462-2  
: Sequence 2, Application US/099888462

Publication No. US20030046726A1  
 GENERAL INFORMATION:  
 APPLICANT: Koziel, Michael G.  
 Desai, Mallini M.  
 Lewis, Kelly S.  
 Kramer, Vance C.  
 Warren, Gregory W.  
 Evola, Stephen V.  
 Crossland, Lyle D.  
 Wright, Martha S.  
 Merlin, Ellis J.  
 Launis, Karen L.  
 TITLE OF INVENTION: SYNTHETIC DNA SEQUENCE HAVING ENHANCED  
 INSECTICIDAL ACTIVITY IN MAIZE  
 NUMBER OF SEQUENCE: 94  
 CORRESPONDENCE ADDRESSES:  
 ADDRESSEE: Syngenta Biotechnology, Inc.  
 STREET: 3054 Cornwallis Road  
 CITY: Research Triangle Park  
 STATE: NC  
 COUNTRY: USA  
 ZIP: 27709  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: PatentIn Release #1.0, Version #1.30  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US 09/988,462  
 FILING DATE: 20-NO. US20030046726A1-2001  
 CLASSIFICATION: <Unknown>  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 09/547,422  
 FILING DATE: 11-APR-2000  
 APPLICATION NUMBER: US 08/459,504  
 FILING DATE: 02-JUN-1995  
 APPLICATION NUMBER: US 07/951,715  
 FILING DATE: 25-SEP-1992  
 APPLICATION NUMBER: US 07/772,027  
 FILING DATE: 04-OCT-1991  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Meigs, J. Timothy  
 REGISTRATION NUMBER: 38,241  
 REFERENCE/DOCKET NUMBER: S-188051  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (919)541-8587  
 TELEFAX: (919)541-8689  
 INFORMATION FOR SEQ ID NO: 2:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 3468 base pairs  
 TYPE: nucleic acid  
 STRANDEDNESS: single  
 TOPOLOGY: linear  
 MOLECULE TYPE: other nucleic acid  
 DESCRIPTION: /desc = "Synthetic DNA"  
 HYPOTHETICAL: NO  
 FEATURE:  
 NAME/KEY: misc\_feature  
 LOCATION: 1..3468  
 OTHER INFORMATION: /product= "Full-length pure maize  
 optimized synthetic Bt"  
 /note= "Disclosed in Figure 3 as syn1T.mze"  
 SEQUENCE DESCRIPTION: SEQ ID NO: 2:  
 US-09-988-462-2

Query Match 3.4%; Score 45.4; DB 9; Length 3468;  
 Best Local Similarity 44.5%; Pred. No. 0.0012;  
 Matches 230; Conservative 0; Mismatches 281; Indels 6; Gaps 1;

QY 251 TCCTCAACAACACGGTGGAGCGGACGCTGACAGCTGAGGCGCTCTCTACACAAG 310  
 DB 2657 TCGTGAACGACGACGACGCGCTGACGAGCGGACGACACAATCGCATGATCCAGCGG 2716

QY 311 TCCTCAACGCGCTGTCTGAGAGCGGCGCTTACTGATGATTGACATGCAACTTTCGCC 370  
 DB 2717 CCACACAAGCGCGTGCACAGCATCCGCGAGGCTTACTGCGCGAGGTGAGGTATCCCG 2776  
 QY 371 GCTACACAGCGCGGCTATCATGCGCCAGGAGCGGTGTGGACGACATCTTTGACCTCT 430  
 DB 2777 GCGTGAACGCGCGGCTATCTTCGAGAGCTGAGAGGCGCGCATCTTCACCGCCTTACGCTGT 2836  
 QY 431 GGGTCCAGATCGCAAAAGTACTAGAGACACACACAAGATCATTTGGCTCATGACG 490  
 DB 2837 ACAGACCGCGGCAAGTATCAAGAACGCGGCTTCAACACGCGCTGAGTGTGGAACG 2896  
 QY 491 AGCGGACGACCTCGACATGATGATCTGGGCGAGAGCTGCAAAAGGTGTCATGCGGA 550  
 DB 2897 TGAAGGCGCACGTGGACGTGGAGAGACAGAACACACCGACCGTGTGTGTCGCCG 2956  
 QY 551 TCCGAAGCGCGGCGGCGCACCTCGACATGATCTCTGCGCGGACCACTTTGCGACG 610  
 DB 2957 AGTGGAGGCGGAGGTGAGGACGAGAGGTGCGGTGTGCGCGCGGCGGTATCATCTGC 3016  
 QY 611 TCGAGACGTATGTCTCCATCGGCGGAGCGGAGACCTTCGCAAGATTACGACCGGATG 670  
 DB 3017 GCGTGACCGGCTTCAAGAGAGGCTTACGGCGAGGCTGCGGACCATCCAGATCGAGA 3076  
 QY 671 GAAGCAGCGATTTGCTGTACTTTGAT-----GTCCACAGTATTCGACATCAACT 724  
 DB 3077 ACACACCGACGACGCTGAAAGTTACGACATCGCTGGAGGAGGAGGTGTACCCCAACACA 3136  
 QY 725 CCGGTCGACGCGGAGTGCACACACAGACAACGTGCA 761  
 DB 3137 CCGTGACCTGCACAGACTACACCGCCACACGAGAGA 3173

RESULT 14  
 US-09-910-186A-9  
 Sequence 9, Application US/09910186A  
 Publication No. US2003009025A1  
 GENERAL INFORMATION:  
 APPLICANT: U.S. Army Medical Research & Materiel Command  
 TITLE OF INVENTION: RECOMBINANT VACCINE AGAINST BOTULINUM  
 FILE REFERENCE: A33626-A 067252.0107  
 CURRENT APPLICATION NUMBER: US/09/910,186A  
 PRIOR FILING DATE: 2001-07-20  
 PRIOR APPLICATION NUMBER: PCT/US00/12890  
 PRIOR FILING DATE: 2000-05-12  
 PRIOR APPLICATION NUMBER: 09/611,419  
 PRIOR FILING DATE: 2000-07-06  
 PRIOR APPLICATION NUMBER: 60/133,865  
 PRIOR FILING DATE: 1999-05-12  
 PRIOR APPLICATION NUMBER: 60/133,866  
 PRIOR FILING DATE: 1999-05-12  
 PRIOR APPLICATION NUMBER: 60/133,867  
 PRIOR FILING DATE: 1999-05-12  
 PRIOR APPLICATION NUMBER: 60/133,868  
 PRIOR FILING DATE: 1999-05-12  
 PRIOR APPLICATION NUMBER: 60/133,869  
 PRIOR FILING DATE: 1999-05-12  
 PRIOR APPLICATION NUMBER: 60/133,873  
 PRIOR FILING DATE: 1999-05-12  
 PRIOR APPLICATION NUMBER: 08/123,975  
 PRIOR FILING DATE: 1993-09-21  
 NUMBER OF SEQ ID NOS: 34  
 SOFTWARE: FastSeq for Windows Version 4.0  
 SEQ ID NO 9  
 LENGTH: 1371  
 TYPE: DNA  
 ORGANISM: Artificial Sequence  
 FEATURE:  
 OTHER INFORMATION: Synthetic construct based on Clostridium botulinum  
 NAME/KEY: CDS  
 LOCATION: (10)...(1359)



CC STRAIN-H37Rv;  
 RX MEDLINE-9825987; PubMed-6634230;  
 RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,  
 RA Gordon S.V., Eiglmeyer K., Gas S., Barry C.E. III, Tekala F.,  
 RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,  
 RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,  
 RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,  
 RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,  
 RA Rafter S., Seeger K., Skelton S., Squares S., Squares R.,  
 RA Sultson J.E., Taylor K., Whitehead S., Barrell B.G.;  
 RA "Deciphering the biology of Mycobacterium tuberculosis from the  
 RT complete genome sequence.";  
 RL Nature 393:537-544(1998).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-CDC 1551 / Oshkosh;  
 RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,  
 RA Peterson J., DeBoy R., Dodson R., Gwin M.L., Hatt D., Hickey E.,  
 RA Kolony J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,  
 RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,  
 RA Bishai W.;  
 RA "Whole genome comparison of Mycobacterium tuberculosis clinical and  
 RT laboratory strains.";  
 RL Submitted (Apr-2003) to the EMBL/GenBank/DBJ databases.  
 CC -1- FUNCTION: Catalyzes the formation of alpha-ketobutyrate from  
 CC threonine in a two-step reaction. The first step is a dehydration  
 CC of threonine, followed by rehydration and liberation of ammonia.  
 CC -1- CATALYTIC ACTIVITY: L-threonine + H(2)O -> 2-oxobutanoate + NH(3) +  
 CC H(2)O.  
 CC -1- COFACTOR: PYRIDOXAL PHOSPHATE (BY SIMILARITY).  
 CC -1- PATHWAY: Isoleucine biosynthesis; first step.  
 CC -1- SIMILARITY: BELONGS TO THE SERINE/THREONINE DEHYDRATASE FAMILY.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: Z74020; CA98332.1; -;  
 DR EMBL: AE007027; AA645877.1; -;  
 DR HSP: P04968; ITDJ.  
 DR TIGR: M1610; -;  
 DR Tuberculist: RV1559; -;  
 DR InterPro: IPR001926; B6\_enzyme\_beta.  
 DR InterPro: IPR000634; S/T\_dehydratase.  
 DR InterPro: IPR001721; ThrDh\_C.  
 DR Pfam: PF00291; PALP; 1.  
 DR Pfam: PF00385; Thr\_dehydrat\_C; 1.  
 DR PROSITE: PS00165; DEHYDRATASE\_SER\_THR; 1.  
 DR KMW: Isolation biosynthesis; Lyase; Pyridoxal phosphate;  
 KM Complete proteome.  
 FT BINDING 66  
 FT DOMAIN 145  
 FT DOMAIN 196  
 FT SEQUENCE 429 AA; 45041 MW; D0C761EC258AC521 CRC64;  
 SO  
 Query Match 50.6%; Score 42; DB 1; Length 429;  
 Best Local Similarity 55.6%; Pred. No. 55;  
 Matches 10; Conservative 3; Mismatches 5; Indels 0; Gaps 0;  
 Oy 2 RATSILAAALAVAGDALA 19  
 Db 254 RAGTLTYAALAAAGDWS 271  
 RESULT 11  
 LY6D\_HUMAN STANDARD; PRT; 128 AA.  
 AC 014210; 092933; 043783;  
 DT 01-NOV-1997 (Rel. 35, Created)

DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DE 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Lymphocyte antigen Ly-6D precursor (E48 antigen).  
 GN LY6D OR E48.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A. AND SEQUENCE OF 21-33.  
 RX MEDLINE-95310346; PubMed-7790363;  
 RA Brakenhoff R.H., van Dijk M., Rood-Knipfels E.M.C., van Dijk M.,  
 RA van Essen H., Weghuis D.O., Sinke R.J., Snow G.B.,  
 RA van Dongen G.A.M.S.;  
 RT "The human E48 antigen, highly homologous to the murine Ly-6 antigen  
 RT TnB, is a GPI-anchored molecule apparently involved in keratinocyte  
 RT cell-cell adhesion.";  
 RL J. Cell Biol. 129:1677-1689(1995).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE-98031741; PubMed-9366413;  
 RA Brakenhoff R.H., van Dijk M., Rood-Knipfels E.M.C., Snow G.B.;  
 RT "A gain of novel tissue specificity in the human Ly-6 gene E48.";  
 RL J. Immunol. 159:4879-4886(1997).  
 RN [3]  
 RP SEQUENCE OF 18-96 FROM N.A.  
 RA Shan X., Bourdeau A., Rhoton A., Wells D.E., Cohen E.H.,  
 RA Landgraf B.E., Palfrey R.G.E.;  
 RL Submitted (SEP-1996) to the EMBL/GenBank/DBJ databases.  
 CC -1- FUNCTION: MAY BE INVOLVED IN CELL-CELL ADHESION AND SIGNAL  
 CC TRANSDUCTION.  
 CC -1- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor.  
 CC -1- TISSUE SPECIFICITY: EXPRESSED EXCLUSIVELY AT THE OUTER CELL  
 CC SURFACE OF TRANSITIONAL EPITHELIA AND THE KERATINOCYTE OF  
 CC STRATIFIED SQUAMOUS EPITHELIA.  
 CC -1- SIMILARITY: CONTAINS 1 UPAR/LY6 DOMAIN.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: X82693; CA58014.1; -;  
 DR EMBL: Y12642; CA73189.1; -;  
 DR EMBL: U66837; AAB07524.1; -;  
 DR InterPro: IPR001526; LY6\_UPAR.  
 DR InterPro: IPR003632; LY-6\_CD59.  
 DR Pfam: PF00021; UPAR\_LY6; 1.  
 DR SMART: SM00134; LU; 1.  
 DR PROSITE: PS00983; LY6\_UPAR; 1.  
 DR T-cell; signal; Antigen; Multigene family; GPI-anchor; Cell adhesion.  
 KW SIGNAL 1  
 FT CHAIN 21  
 FT CHAIN 798  
 FT PROPEP 299  
 FT DOMAIN 21  
 FT DOMAIN 108  
 FT DISULFID 23  
 FT DISULFID 45  
 FT DISULFID 26  
 FT DISULFID 32  
 FT DISULFID 38  
 FT DISULFID 63  
 FT DISULFID 67  
 FT DISULFID 86  
 FT DISULFID 87  
 FT DISULFID 92  
 FT LIPID 98  
 FT LIPID 98  
 FT CONFLICT 10  
 FT CONFLICT 10  
 FT CONFLICT 60  
 FT CONFLICT 60  
 FT CONFLICT 76  
 FT CONFLICT 76  
 FT SEQUENCE 128 AA; 13286 MW; 39618DF6AB5B0EBD CRC64;  
 SO  
 Query Match 50.0%; Score 41.5; DB 1; Length 128;  
 Best Local Similarity 68.4%; Pred. No. 24;  
 Matches 13; Conservative 0; Mismatches 5; Indels 1; Gaps 1;

```

-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: U05673; AAA19001.1; -
CC HSSP: P29274; 1MMH.
CC MGD: MGI:99403; Adora2b.
CC InterPro: IPR000276; GPCR_Rhodopsn.
CC Pfam: PF00001; 7tm_1; 1.
CC PRINTS: PR00237; GPCRHHODPSN.
CC PROSITE: PS00237; G_PROTEIN_RECEPT_FL_1; 1.
CC PROSITE: PS50262; G_PROTEIN_RECEPT_FL_2; 1.
CC G-protein coupled receptor; Transmembrane; Glycoprotein;
CC Lipoprotein; Palmitate.
CC
CC FT DOMAIN 1 8 EXTRACELLULAR (POTENTIAL).
CC FT TRANSMEM 9 31 1 (POTENTIAL).
CC FT DOMAIN 32 44 CYTOPLASMIC (POTENTIAL).
CC FT TRANSMEM 45 67 2 (POTENTIAL).
CC FT DOMAIN 68 78 EXTRACELLULAR (POTENTIAL).
CC FT TRANSMEM 79 101 3 (POTENTIAL).
CC FT DOMAIN 102 121 CYTOPLASMIC (POTENTIAL).
CC FT TRANSMEM 122 144 4 (POTENTIAL).
CC FT DOMAIN 145 178 EXTRACELLULAR (POTENTIAL).
CC FT TRANSMEM 179 203 5 (POTENTIAL).
CC FT DOMAIN 204 235 CYTOPLASMIC (POTENTIAL).
CC FT TRANSMEM 236 259 6 (POTENTIAL).
CC FT DOMAIN 260 267 EXTRACELLULAR (POTENTIAL).
CC FT TRANSMEM 268 291 7 (POTENTIAL).
CC FT DOMAIN 292 332 CYTOPLASMIC (POTENTIAL).
CC FT TRANSMEM 331 311 BY SIMILARITY.
CC FT DISULFID 78 171 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC FT CARBOHYD 153 153 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC FT CARBOHYD 163 163 PALMITATE (POTENTIAL).
CC FT LIPID 311 311 PALMITATE (POTENTIAL).
CC SO SEQUENCE 332 AA; 36064 MW; CBA3108371AD1C1C CRC64;
Query Match 50.6%; Score 42; DB 1; Length 332;
Best Local Similarity 62.5%; Pred. No. 45;
Matches 10; Conservative 2; Mismatches 4; Indels 0; Gaps 0;
OY 3 ATSLAALAVAGDAL 18
DB 12 ALELVIALAVAGNVL 27
RESULT 9
AA2B_RAT STANDARD: PRT; 332 AA.
ID AA2B_RAT
AC P29276;
DT 01-DEC-1992 (Rel. 24, Created)
DT 01-DEC-1992 (Rel. 24, Last sequence update)
DT 01-FEB-1996 (Rel. 33, Last annotation update)
DE Adenosine A2b receptor.
GN ADORA2B.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN (1)
RP SEQUENCE FROM N.A.
RX TISSUE=Brain;
RX MEDLINE=92261592; PubMed=1584214;
RX Stehle J.H., Rivkees S.A., Lee J.J., Weaver D.R., Deeds J.D.,
RX Reppert S.M.;
RX "Molecular cloning and expression of the cDNA for a novel
RX A2-adenosine receptor subtype.";
RX Mol. Endocrinol. 6:384-393(1992).
RN (1)
RP CHARACTERIZATION.

```

```

RX MEDLINE=93078793; PubMed=1333049;
RX Rivkees S.A., Reppert S.M.;
RX "REF1 encodes an A2b-adenosine receptor.";
RX Mol. Endocrinol. 6:1598-1604(1992).
CC -1- FUNCTION: RECEPTOR FOR ADENOSINE. THE ACTIVITY OF THIS RECEPTOR
CC IS MEDIATED BY G PROTEINS WHICH ACTIVATE ADENYLYL CYCLASE.
CC -1- SUBCELLULAR LOCATION: Integral membrane protein.
CC -1- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: M91466; AAA20981.1; -
CC HSSP: P29274; 1MMH.
CC PIR: A42171; A42171.
CC InterPro: IPR000276; GPCR_Rhodopsn.
CC Pfam: PF00001; 7tm_1; 1.
CC PRINTS: PR00237; GPCRHHODPSN.
CC PROSITE: PS00237; G_PROTEIN_RECEPT_FL_1; 1.
CC PROSITE: PS50262; G_PROTEIN_RECEPT_FL_2; 1.
CC G-protein coupled receptor; Transmembrane; Glycoprotein;
CC Lipoprotein; Palmitate.
CC
CC FT DOMAIN 1 8 EXTRACELLULAR (POTENTIAL).
CC FT TRANSMEM 9 31 1 (POTENTIAL).
CC FT DOMAIN 32 44 CYTOPLASMIC (POTENTIAL).
CC FT TRANSMEM 45 67 2 (POTENTIAL).
CC FT DOMAIN 68 78 EXTRACELLULAR (POTENTIAL).
CC FT TRANSMEM 79 101 3 (POTENTIAL).
CC FT DOMAIN 102 121 CYTOPLASMIC (POTENTIAL).
CC FT TRANSMEM 122 144 4 (POTENTIAL).
CC FT DOMAIN 145 178 EXTRACELLULAR (POTENTIAL).
CC FT TRANSMEM 179 203 5 (POTENTIAL).
CC FT DOMAIN 204 235 CYTOPLASMIC (POTENTIAL).
CC FT TRANSMEM 236 259 6 (POTENTIAL).
CC FT DOMAIN 260 267 EXTRACELLULAR (POTENTIAL).
CC FT TRANSMEM 268 291 7 (POTENTIAL).
CC FT DOMAIN 292 332 CYTOPLASMIC (POTENTIAL).
CC FT DISULFID 78 171 BY SIMILARITY.
CC FT CARBOHYD 153 153 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC FT CARBOHYD 163 163 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC FT LIPID 311 311 PALMITATE (POTENTIAL).
CC SO SEQUENCE 332 AA; 36367 MW; FOABABSCCA858 CRC64;
Query Match 50.6%; Score 42; DB 1; Length 332;
Best Local Similarity 62.5%; Pred. No. 45;
Matches 10; Conservative 2; Mismatches 4; Indels 0; Gaps 0;
OY 3 ATSLAALAVAGDAL 18
DB 12 ALELVIALAVAGNVL 27
RESULT 10
THD1_MYCTU STANDARD: PRT; 429 AA.
ID THD1_MYCTU
AC Q10766;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Probable threonine dehydrogenase biosynthetic (EC 4.2.1.16) (Threonine
DE deaminase).
GN IIVA OR RV1559 OR MT1610 OR MTCY48.06C.
OS Mycobacterium tuberculosis.
OC Bacteria; Actinobacteria; Actinobacteria (class); Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN (1)
RP SEQUENCE FROM N.A.

```

```

RT human tumors."
RL Genomics 17:762-764(1993).
RN [3]
RP SEQUENCE FROM N.A.
RC TISSUE-Muscle;
RA Strausberg R.;
RL submitted (Aug-2001) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: PROHIBITIN INHIBITS DNA SYNTHESIS. IT HAS A ROLE IN
CC REGULATING PROLIFERATION. AS YET IT IS UNCLEAR IF THE PROTEIN
CC OR THE MRNA EXHIBITS THIS EFFECT.
CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
CC -1- TISSUE SPECIFICITY: WIDELY EXPRESSED IN DIFFERENT TISSUES.
CC -1- DISEASE: MUTATED IN SPORADIC BREAST CANCER.
CC -1- SIMILARITY: BELONGS TO THE PROHIBITIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; S85655; AAB21614.1; -
DR EMBL; L14272; -; NOT ANNOTATED_CDS.
DR EMBL; BC013401; AAH13401.1; -
DR SWISS-2DPAGE; P35232; HUMAN.
DR PMAA-2DPAGE; P35232; -
DR Genew; HGNC:8912; PHB.
DR MIM; 176705; -
DR InterPro: IPR001107; Band_7.
DR InterPro: IPR000163; Prohibitin.
DR Pfam: PF01145; Band_7.1.
DR PRINTS; PR00679; PROHIBITIN.
DR SMART; SM00244; PHB; 1.
DR DNA synthesis; Disease mutation; Proto-oncogene.
FT VARIANT 88 88 V->A (IN BREAST CANCER).
FT FT FT R->H (IN BREAST CANCER).
FT VARIANT 105 105 /FTID-VAR_006479.
FT FT FT /FTID-VAR_006480.
SQ SEQUENCE 272 AA; 29804 MW; 915494273E342C76 CRC64;

Query Match 51.8%; Score 43; DB 1; Length 272;
Best Local Similarity 52.9%; Pred. No. 27;
Matches 9; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

OY 2 RATSLLAALAVAGDAL 18
Db 219 KAAELIANSLATAGDGL 235

RESULT 7
PHB_MOUSE
ID PHB_MOUSE STANDARD; PRT; 272 AA.
AC P24142;
DT 01-MAR-1992 (Rel. 21, Last Created)
DT 01-MAR-1992 (Rel. 21, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE Prohibitin (B-cell receptor associated protein 32) (BAP 32).
GN PHB.
OS Mus musculus (Mouse), and
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090, 10116;
RN (1)
RP SEQUENCE FROM N.A.
RC SPECIES-Rat;
RX MEDLINE-91141485; PubMed-1996099;
RA Nuehl M.J., Stewart D.A., Walker L., Friedman V., Wood C.M.,
RA Owens G.A., Smith J.R., Schneider E.L., Dell'Orco R., Lumpkin C.K.,
RA Danner D.B., McClung J.K.;
*Prohibitin, an evolutionarily conserved intracellular protein that

```

```

RT blocks DNA synthesis in normal fibroblasts and HeLa cells."
RL Mol. Cell. Biol. 11:1372-1381(1991).
RN [2]
RP SEQUENCE FROM N.A.
RC SPECIES-Mouse; STRAIN-BALB/C; TISSUE-Lymphoid;
RX MEDLINE-94349926; PubMed-8070406;
RA Terashima M., Kim K.M., Adachi T., Nielsen P.J., Reth M.,
RA Koehler G., Lamers M.C.;
RT "The IgM antigen receptor of B lymphocytes is associated with
RT prohibitin and a prohibitin-related protein."
RL EMBO J. 13:3782-3792(1994).
CC -1- FUNCTION: PROHIBITIN INHIBITS DNA SYNTHESIS. IT HAS A ROLE IN
CC REGULATING PROLIFERATION. AS YET IT IS UNCLEAR IF THE PROTEIN
CC OR THE MRNA EXHIBITS THIS EFFECT.
CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
CC -1- TISSUE SPECIFICITY: WIDELY EXPRESSED IN DIFFERENT TISSUES.
CC -1- SIMILARITY: BELONGS TO THE PROHIBITIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M61219; AAA63500.1; -
DR EMBL; X78682; CA55349.1; -
DR PIR; A39682; A39682.
DR SWISS-2DPAGE; P24142; MOUSE.
DR MGI; MGI:97572; PHB.
DR InterPro: IPR001107; Band_7.
DR InterPro: IPR000163; Prohibitin.
DR Pfam; PF01145; Band_7.1.
DR PRINTS; PR00679; PROHIBITIN.
DR SMART; SM00244; PHB; 1.
DR DNA synthesis.
KW K W
SQ SEQUENCE 272 AA; 29820 MW; 6B26073E169C2FEC CRC64;

Query Match 51.8%; Score 43; DB 1; Length 272;
Best Local Similarity 52.9%; Pred. No. 27;
Matches 9; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

OY 2 RATSLLAALAVAGDAL 18
Db 219 KAAELIANSLATAGDGL 235

RESULT 8
AAZB_MOUSE
ID AAZB_MOUSE STANDARD; PRT; 332 AA.
AC O60614;
DT 01-NOV-1997 (Rel. 35, Last Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE Adenosine A2b receptor.
GN ADORA2B.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN-BALB/C; TISSUE-Bone marrow;
RX MEDLINE-94209670; PubMed-8157966;
RA Marguardt D.L., Walker L.L., Heinemann S.;
RT "Cloning of two adenosine receptor subtypes from mouse bone marrow-
RT derived mast cells."
RL J. Immunol. 152:4508-4515(1994).
CC -1- FUNCTION: RECEPTOR FOR ADENOSINE. THE ACTIVITY OF THIS RECEPTOR
CC IS MEDIATED BY G PROTEINS WHICH ACTIVATE ADENYLYL CYCLASE.
CC -1- SUBCELLULAR LOCATION: Integral membrane protein.
CC -1- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.

```

RT "The cDNA and derived amino acid sequences of human and bovine bone  
RT Gla protein.";  
RL Nucleic Acids Res. 18:1909-1909(1990).  
RN (2)  
RP SEQUENCE FROM N.A.  
RX MEDLINE-87004555; PubMed-3019668;  
RA Celeste A.J., Buecker J.L., Kriz R., Wang E.A., Wozney J.M.;  
RT "Isolation of the human gene for bone gla protein utilizing mouse and  
RT rat cDNA clones";  
RL EMBO J. 5:1885-1890(1986).  
RN [3]  
RP SEQUENCE OF 52-100.  
RX MEDLINE-81006914; PubMed-6967872;  
RA Poser J.W., Esch F.S., Ling N.C., Price P.A.;  
RT "Isolation and sequence of the vitamin K-dependent protein from human  
RT bone. Undercarboxylation of the first glutamic acid residue.";  
RL J. Biol. Chem. 255:8685-8691(1980).  
CC -1- FUNCTION: CONSTITUTES 1-2% OF THE TOTAL BONE PROTEIN. IT BINDS  
CC STRONGLY TO APATITE AND CALCIUM.  
CC -1- PTM: GAMMA-CARBOXYGLUTAMIC ACID RESIDUES ARE FORMED BY VITAMIN K  
CC DEPENDENT CARBOXYLATION. THESE RESIDUES ARE ESSENTIAL FOR THE  
CC BINDING OF CALCIUM.  
CC -1- SIMILARITY: BELONGS TO THE OSTEOCALCIN / MATRIX GLA-PROTEIN  
CC FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
CC -----  
DR EMBL: X53698; CAA3736.1; -  
DR EMBL: X51699; CAA3596.1; -  
DR EMBL: X04143; CAA27763.1; -  
DR PIR: A03301; GEHU.  
DR PIR: G25471; C25471.  
DR PIR: S12652; S12652.  
DR Genew: HGNC:1043; BGLAP.  
DR MIM: 112260; -  
DR InterPro: IPR002384; GLA\_bone.  
DR InterPro: IPR00294; VICK\_dep\_GLA.  
DR Pfam: PF00594; gla; 1.  
DR PRINTS: PR00002; GLABONE.  
DR SMART: SM00069; GLA; 1.  
DR PROSITE: PS00011; GLU-CARBOXYLATION; 1.  
KM Calcium-binding; Gamma-carboxyglutamic acid; Vitamin K; Bone; Signal.  
FT SIGNAL 1 23  
FT PROPEP 24 51 PROBABLE.  
FT CHAIN 52 100 OSTEOCALCIN.  
FT MOD\_RES 68 68 GAMMA-CARBOXYGLUTAMIC ACID (IN 9% OF THE  
FT MOD\_RES 72 72 MOLECULES).  
FT MOD\_RES 75 75 GAMMA-CARBOXYGLUTAMIC ACID.  
FT DISULFID 74 80 GAMMA-CARBOXYGLUTAMIC ACID.  
FT CONFLICT 33 34 MISSING (IN REF. 2).  
SQ SEQUENCE 100 AA; 10962 MW; 4DF2A0A80849CB71 CRC64;

Query Match 51.8%; Score 43; DB 1; Length 100;  
Best Local Similarity 57.1%; Pred. No. 12;  
Matches 12; Conservative 2; Mismatches 3; Indels 4; Gaps 1;

QY 1 MRATSLA-----AALAVAGDA 17  
DB 1 MRATLALLAALALCIAGQA 21

RESULT 5  
COMB\_ANASP STANDARD; PRT; 245 AA.  
AC Q8Y125;  
DT 15-JUN-2002 (Rel. 41, Created)

DT 15-JUN-2002 (Rel. 41, Last sequence update)  
DE 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Probable 2-phosphosulfolactate phosphatase (EC 3.1.3.-).  
GN COMB OR ALI2568.  
OS Anabaena sp. (strain PCC 7120).  
OC Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Nostoc.  
ON NCBI\_TaxID=103690;  
RX (1)  
RP SEQUENCE FROM N.A.  
RX MEDLINE-21595285; PubMed-11759840;  
RA Kaneko T., Nakamura Y., Wolk C.P., Kuritz T., Sasamoto S.,  
RA Watanabe A., Iiguchi M., Ishikawa A., Kawashima K., Kimura T.,  
RA Kishida Y., Kohara M., Matsumoto M., Matsuno A., Muraki A.,  
RA Nakazaki N., Shimpo S., Sugimoto M., Takazawa M., Yamada M.,  
RA Yasuda M., Tabata S.;  
RT "Complete genomic sequence of the filamentous nitrogen-fixing  
RT cyanobacterium Anabaena sp. strain PCC 7120.";  
RL DNA Res. 8:205-213(2001).  
CC -1- CATALYTIC ACTIVITY: 2-phosphosulfolactate - sulfolactate +  
CC phosphate.  
CC -1- COFACTOR: Magnesium (By similarity).  
CC -1- SIMILARITY: BELONGS TO THE COMB FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
CC -----  
DR EMBL: AP003589; BAB74267.1; -  
DR InterPro: IPR005238; 2-ph-phosphatase.  
DR TIGRFAMs: TIGR00298; TIGR00298; 1.  
KM Hydrolyase; Magnesium; Complete proteome.  
SQ SEQUENCE 245 AA; 26694 MW; BC7F48BA4E5B658 CRC64;

Query Match 51.8%; Score 43; DB 1; Length 245;  
Best Local Similarity 44.4%; Pred. No. 25;  
Matches 8; Conservative 5; Mismatches 5; Indels 0; Gaps 0;

QY 1 MRATSLAALAVAGDAL 18  
DB 27 LRATSTIATVLSAGEAV 44

RESULT 6  
PHB\_HUMAN STANDARD; PRT; 272 AA.  
AC P35232;  
DT 01-FEB-1994 (Rel. 28, Created)  
DT 01-FEB-1994 (Rel. 28, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Prohibitin.  
GN PHB.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
ON NCBI\_TaxID=9606;  
RX (1)  
RP SEQUENCE FROM N.A.  
RX TISSUE=Breast;  
RC MEDLINE-92174193; PubMed-1540973;  
RA Sato T., Saito H., Swensen J., Olfant A., Wood C., Danner D.,  
RA Sakamoto T., Takita K., Kasumi F., Miki Y., Skolnick M., Nakamura Y.;  
RT "The human prohibitin gene located on chromosome 17q21 is mutated in  
RT sporadic breast cancer.";  
RL Cancer Res. 52:1643-1646(1992).  
RN [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE-94063925; PubMed-8244394;  
RA Sato T., Sakamoto T., Takita K., Saito H., Okui K., Nakamura Y.;  
RT "The human prohibitin (PHB) gene family and its somatic mutations in

OY 1 MRATSLAALAVAGDALA 19  
 DB 5 LRMVLAALAAVEGVAGNALA 23

## RESULT 2

Y13L\_HAEIN STANDARD; PRT; 346 AA.  
 AC P43951;  
 DT 01-NOV-1995 (Rel. 32, Created)  
 DT 01-NOV-1995 (Rel. 32, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Protein HI0131 precursor.  
 GN HI0131.  
 OS Haemophilus influenzae.  
 OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;  
 OC Haemophilus  
 OC NCBI\_TaxID=127;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=RD / KW20 / ATCC 51907;  
 RX MEDLINE=95350630; PubMed=7542800;  
 RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,  
 Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,  
 McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,  
 Scott J.D., Shiley R., Liu L.-I., Glodek A., Kelley J.M.,  
 Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,  
 Uterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,  
 Fine L.D., Fritchman J.L., Fuhrmann J.L., Geoghegan N.S.M.,  
 Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,  
 Venter J.C.;  
 RT "Whole-genome random sequencing and assembly of Haemophilus influenzae  
 Rd.";  
 RL Science 269:496-512(1995).

RN [2]  
 RP IDENTIFICATION BY MASS SPECTROMETRY.  
 RX MEDLINE=20137488; PubMed=10675023;  
 RA Langen H., Takacs B., Evers S., Berndt P., Lahm H.W., Wipf B.,  
 Gray C., Fountoulakis M.;  
 RT "Two-dimensional map of the proteome of Haemophilus influenzae.";  
 RL Electrophoresis 21:411-429(2000).

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL; U32699; AAC21803.1; -  
 DR TIGR: HI0131;  
 DR InterPro: IPR000557; SBP\_bac\_1.  
 DR Pfam: PF01547; SBP\_bacterial\_1; 1.  
 KW Signal; Complete proteome.  
 FT SIGNAL 1 27  
 FT CHAIN 28 346 POTENTIAL.  
 FT SEQUENCE 346 AA; 37898 MW; 233AD90C54E10710 CRC64;

Query Match 56.68; Score 47; DB 1; Length 346;  
 Best Local Similarity 64.7%; Pred. No. 9.2;  
 Matches 11; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

OY 3 ATSLAALAVAGDALA 19  
 DB 11 STALLAAGLAAGSANA 27

## RESULT 3

XYN2\_MAGGR STANDARD; PRT; 233 AA.  
 ID XYN2\_MAGGR  
 AC P55335; Q01171;

DT 01-OCT-1996 (Rel. 34, Created)  
 DT 01-OCT-1996 (Rel. 34, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Endo-1,4-beta-xylanase 22 precursor (EC 3.2.1.8) (Xylanase 22)  
 DE (1,4-beta-D-xylan xylanohydrolase 22).  
 GN XYN22.

OS Magnaporthe grisea (Rice blast fungus) (Pyricularia grisea);  
 OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;  
 OC Sordariomycetes incertae sedis; Magnaportheaceae; Magnaporthe  
 OC NCBI\_TaxID=148305;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=KEN60-19;  
 RX MEDLINE=96172742; PubMed=8589407;  
 RA Wu S.C., Kaufman S., Davyll A.G., Albersheim P.;

RT "Purification, cloning and characterization of two xylanases from  
 RT Magnaporthe grisea, the rice blast fungus.";  
 RL Mol. Plant Microbe Interact. 8:506-514(1995).  
 CC -1- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-xylosidic  
 CC linkages in xylans.  
 CC -1- PATHWAY: Xylan degradation.  
 CC -1- SIMILARITY: BELONGS TO CELLULASE FAMILY G (FAMILY 11 OF GLYCOSYL  
 CC HYDROLASES).

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL; L37529; AAC41683.1; -  
 DR HSSP: O43097; IYNA.  
 DR InterPro: IPR001137; GH\_11.  
 DR Pfam: PF00457; Glyco\_hydro\_11; 1.  
 DR PRINTS: PR00911; GLHYDRLASE1.  
 DR PROSITE: PS00776; GLYCOSYL\_HYDROL\_F11\_1; 1.  
 DR PROSITE: PS00777; GLYCOSYL\_HYDROL\_F11\_2; 1.  
 KW Xylan degradation; Hydrolase; Glycosidase; Signal.  
 FT SIGNAL 1 39 POTENTIAL.  
 FT CHAIN 40 233 ENDO-1,4-BETA-XYLANASE 22.  
 FT ACT\_SITE 126 126 NUCLEOPHILE (BY SIMILARITY).  
 FT ACT\_SITE 217 217 PROTON DONOR (BY SIMILARITY).  
 SQ SEQUENCE 233 AA; 25491 MW; 400963836F581F98 CRC64;

Query Match 53.08; Score 44; DB 1; Length 233;  
 Best Local Similarity 56.28; Pred. No. 18;  
 Matches 9; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

OY 4 TSLAALAVAGDALA 19  
 DB 5 TSIYTAVALAAGSALA 20

## RESULT 4

OSTC\_HUMAN STANDARD; PRT; 100 AA.  
 ID OSTC\_HUMAN  
 AC P02818;  
 DT 21-JUL-1986 (Rel. 01, Created)  
 DT 01-AUG-1990 (Rel. 15, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Osteocalcin precursor (Gamma-carboxyglutamic acid-containing protein)  
 DE (bone Gla-protein) (BGP).  
 GN BGLAP.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 OC NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=90245603; PubMed=2336375;  
 RA Kiefer M.C., Saphire A.C.S., Bauer D.M., Barr P.J.;



R:Moore, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, I.  
 erage, G.; Gillet, W.; Grant, C.; Guenther, D.; Kutyavina, T.; Levy, R.; Li, M.; McCell  
 A:Title: Isolation of the human gene for bone gla protein utilizing mouse and rat cDN  
 Science 294, 2317-2323, 2001  
 A:Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm,  
 ster, E.W.  
 A:Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.  
 A:Reference number: AB2577; PMID:11743193  
 A:Accession: A13234  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-320 <KUR>  
 A:Cross-references: GB:AE008690; PIDN:AA46295.1; PID:q17744078; GSPDB:GN00189  
 A:Experimental source: strain C58 (dupont)  
 A:Genetics:  
 A:Gene: ssua  
 A:Genome: plasmid

Query Match 53.0%; Score 44; DB 2; Length 320;  
 Best Local Similarity 64.7%; Pred. No. 31;  
 Matches 11; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

OY 3 ATSLAALAVAGDALA 19  
 |||:||||| |||  
 Db 10 ATALVAAALMAGSAA 26

RESULT 13  
 A11033  
 Hypothetical protein SRY4596 [imported] - Salmonella enterica subsp. enterica serovar ty  
 C:Species: Salmonella enterica subsp. enterica serovar typh  
 A:Note: this species has also been called Salmonella typh  
 C:Date: 09-Nov-2001 #sequence, revision 09-Nov-2001 #text\_change 09-Nov-2001  
 C:Accession: A11033  
 R:Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Churcher,  
 th, T.; Conington, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar,  
 S.; Moule, S.; O'Garra, P.  
 Nature 413, 848-852, 2001  
 A:Authors: Park, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.;  
 A:Title: Complete genome sequence of a multiple drug resistant Salmonella enterica serov  
 A:Reference number: AB0502; PMID:11677608  
 A:Accession: A11033  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-89 <PAR>  
 A:Cross-references: GB:AL513382; PIDN:CAD06718.1; PID:q16505371; GSPDB:GN00176  
 C:Genetics:  
 A:Gene: SRY4596

Query Match 51.8%; Score 43; DB 2; Length 89;  
 Best Local Similarity 57.9%; Pred. No. 14;  
 Matches 11; Conservative 1; Mismatches 7; Indels 0; Gaps 0;

OY 1 MRATSLAALAVAGDALA 19  
 |||:||||| |||  
 Db 1 MRINWLLTSLVAGPALA 19

RESULT 14  
 GEHU  
 osteocalcin precursor [validated] - human  
 M:Alternate names: BGP; bone gla protein; gamma-carboxyglutamic acid-containing protein  
 C:Species: Homo sapiens (man)  
 C:Date: 31-Dec-1980 #sequence, revision 07-Apr-1994 #text\_change 08-Dec-2000  
 C:Accession: S12652; C25471; A03301; S08694  
 R:Kiefer, M.C.; Saphire, A.C.S.; Bauer, D.M.; Barr, P.J.  
 Nucleic Acids Res. 18, 1909, 1990  
 A:Title: The cDNA and derived amino acid sequences of human and bovine bone gla protein.  
 A:Reference number: S12652; MUID:90245603; PMID:2336375  
 A:Accession: S12652  
 A:Molecule type: mRNA  
 A:Residues: 1-100 <KIE>  
 A:Cross-references: EMBL:X53698; NID:q36092; PIDN:CAA37736.1; PID:q36093

R:Celeste, A.J.; Rosen, V.; Buecker, J.L.; Kriz, R.; Wang, E.A.; Wozney, J.M.  
 EMBO J. 5, 1885-1890, 1986  
 A:Title: Isolation of the human gene for bone gla protein utilizing mouse and rat cDN  
 A:Reference number: A91045; MUID:87004555; PMID:3019668  
 A:Accession: C25471  
 A:Molecule type: DNA  
 A:Residues: 1-32,35-100 <CEL>  
 A:Cross-references: EMBL:X04143; NID:q29449; PIDN:CAA27763.1; PID:q29450  
 R:Poser, J.W.; Esch, F.S.; Ling, N.C.; Price, P.A.  
 J. Biol. Chem. 255, 8685-8691, 1980  
 A:Title: Isolation and sequence of the vitamin K-dependent protein from human bone. O  
 A:Reference number: A03301; MUID:81006914; PMID:6967872  
 A:Accession: A03301  
 A:Molecule type: protein  
 A:Residues: 52-100 <POS>  
 R:Calins, J.R.; Williamson, M.K.; Price, P.A.  
 Anal. Biochem. 199, 93-97, 1991  
 A:Title: Direct identification of gamma-carboxyglutamic acid in the sequencing of vit  
 A:Reference number: A44566; MUID:92222128; PMID:1807167  
 A:Contents: annotation  
 C:Comment: This protein, isolated from bone, binds strongly to apatite.  
 C:Comment: Alternative splicing may produce the sequence presented in reference A9104  
 C:Comment: Glu-68 is gamma-carboxylated in 9-50% of the molecules.  
 C:Genetics:  
 A:Gene: GDB:BGLAP  
 A:Cross-references: GDB:118760; OMIM:112260  
 A:Map position: 1q25-1q31  
 A:Introns: 22/1; 35/1; 58/2  
 C:Superfamily: osteocalcin  
 C:Keywords: bone; calcium binding; carboxyglutamic acid; extracellular matrix  
 F:1-51/Domain: signal sequence #status predicted <SIG>  
 F:52-100/Product: osteocalcin #status experimental <MAT>  
 F:60/Modified site: 4-hydroxyproline (Pro) #status absent  
 F:68/Modified site: gamma-carboxyglutamic acid (Glu) #status experimental  
 F:72,75/Modified site: gamma-carboxyglutamic acid (Glu) #status experimental  
 F:74-80/Disulfide bonds: #status experimental

Query Match 51.8%; Score 43; DB 1; Length 100;  
 Best Local Similarity 57.1%; Pred. No. 15;  
 Matches 12; Conservative 2; Mismatches 3; Indels 4; Gaps 1;

OY 1 MRATSLAALAVAGDA 17  
 |||:||||| |||  
 Db 1 MRATLTLALALACIAGQA 21

RESULT 15  
 C84543  
 probable Hesp-like protein [imported] - Arabidopsis thaliana  
 C:Species: Arabidopsis thaliana (mouse-ear cress)  
 C:Date: 02-Feb-2001 #sequence, revision 02-Feb-2001 #text\_change 17-May-2002  
 C:Accession: C84543  
 R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.  
 M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon,  
 euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter  
 Nature 402, 761-768, 1999  
 A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.  
 A:Reference number: AB4420; MUID:20083487; PMID:10617197  
 A:Accession: C84543  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-137 <STO>  
 A:Cross-references: GB:AE002093; NID:q4581114; PIDN:AMD24604.1; GSPDB:GN00139  
 C:Genetics:  
 A:Gene: At2g16710  
 A:Map position: 2  
 C:Superfamily: conserved hypothetical protein HI0376

Query Match 51.8%; Score 43; DB 2; Length 137;  
 Best Local Similarity 55.6%; Pred. No. 20;  
 Matches 10; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

OY 1 MRATSLAALAVAGDAL 18

Genome Res. 11, 731-753, 2001  
A:Title: The complete genome sequence of the lactic acid bacterium *Lactococcus lactis* ssp. *lactis*  
A:Reference number: A86625; MUID:21235186; PMID:11337471  
A:Accession: F86767  
A:Status: Preliminary  
A:Molecule type: DNA  
A:Residues: 1-166 <STO>  
A:Cross-references: GB:AE005176; PID:912724104; PIDN:AAK05240.1; GSPDB:GN00146  
A:Experimental source: strain IL1403  
C:Genetics:  
A:Gene: y1lf

	Query Match	Score	DB	Length
Best Local Similarity	54.2%	45	2	166
Matches	9	Conservative	4	Mismatches
			3	Indels
			0	Gaps
Qy	1	MRATSLAAALVACD	16	
	:	: :		
Db	22	MRISLLVLAQISGE	37	

RESULT 8  
E87300  
conserved hypothetical protein CC0414 [Imported] - Caulobacter crescentus  
C:Species: Caulobacter crescentus  
C:Date: 20-Apr-2001 #sequence:revision 20-Apr-2001 #text:change 20-Apr-2001  
C:Accession: E87300  
R:Norman, W.C.; Feldblyum, T.V.; Paulsen, I.T.; Neilson, K.E.; Eisen, J.; Heidelberg, J.  
B.; Laub, M.T.; DeBoy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Kolon  
N., J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C.M.  
Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001  
A:Title: Complete Genome Sequence of Caulobacter crescentus.  
A:Reference number: A87249; MUID:21173698; PMID:11259647  
A:Accession: E87300  
A:Status: Preliminary  
A:Molecule type: DNA  
A:Residues: 1-178 <SMO>  
A:Cross-references: GB:AE005673; NID:G13421577; PID:AAK22401.1; GSPDB:GN00148  
C:Genetics:  
:Gene: CC0414

Query Match	54.2%	Score 45	DB 2	Length 176
Best Local Similarity	66.7%	Pred. No. 13		
Matches 12	Conservative 0	Mismatches 6	Indels 0	Gaps 0
OY	2	RATSLLAALAVAGDALA	19	
	1	1	1	1
	1	1	1	1
db	8	RAAVALALAAAAGPALA	25	

```

RESULT 9
T34931
hypothetical protein SC3F9.09 SC3F9.09 - Streptomyces coelicolor
C:Species: Streptomyces coelicolor
C:Date: 05-Nov-1999 #sequence_revision 05-Nov-1999 #text_change 05-Nov-1999
C:Accession: T34931
R:Seeger, R.J.; Harris, D.; Parkhill, J.; Barrett, B.G.; Rajandream, M.A.
submitted to the EMBL Data Library, June 1998
A:Reference number: Z21562
A:Accession: T34931
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-363 <SEE>
A:Cross-references: EMBL:AL023862; PIDN:CA119632.1; GSFDB:GN00070; SCOEDB:SC3F9.09
A:Experimental source: strain A3(2)
C:Genetics:
A:Gene: SCOEDB:SC3F9.09

```

Query Match	54.28;	Score 45;	DB 2;	Length 363;
Best Local Similarity	84.68;	Pred. No. 24;		
Matches 11; Conservative	0;	Mismatches 2;	Indels 0;	Gaps 0;
0y 3 ATSLAAATAVAG 15				

```

Db      336  ATSLVVALAPAG  348
          |||||  |||||  ||

```

RESULT 10  
 H70836  
 hypothetical protein RV0290 - Mycobacterium tuberculosis (strain H37RV)  
 C:Species: Mycobacterium tuberculosis  
 C:Date: 17-Jul-1998 #sequence\_revision 17-Jul-1998 #text\_change 22-Oct-1999  
 C:Accession: H70836  
 R:Cole, S.T.; Brosch, R.; Parkhill, J.; Garlier, T.; Churcher, C.; Harris, D.; Gordon  
 J.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holtroyd,  
 Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.  
 Nature 393, 537-544, 1998  
 A:Authors: Scares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrrell, B.G.  
 A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete geno  
 A:Reference number: A70500; MIMD:98293987; PMID:9654250  
 A:Accession: H70836  
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-472 <COL>  
 A:Cross-references: GB:AL021930; GB:AL123456; NID:93261524; PIDN:CA17365.1; PID:el25  
 A:Experimental source: strain H37RV  
 C:Genetics:  
 A:Gene: RV0290

```

Query Match      54.2%  Score 45:  DB 2:  Length 472:
Best Local Similarity 64.7%:
Matches 11:  Conservative 2:  Mismatches 4:  Indels 0:  Gaps 0:

QY      3  ATSLAAALAVAGDALA 19
      11:11 11111 1:1
Db      151 ATGYLAGLAVAGTAA 167

```

RESULT 11  
 E82235  
 hypothetical protein VC1154 [Imported] - *Vibrio cholerae* (strain N16961 serogroup O1)  
 C:Species: *Vibrio cholerae*  
 C:Date: 18-Aug-2000 #sequence\_revision 20-Aug-2000 #text\_change 02-Feb-2001  
 C:Accession: E82235  
 R:Heidelberg, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwinn, M.L.; Dodson, R.  
 charidson, D.; Ermolaeva, M.D.; Vamathevan, J.; Bass, S.; Qin, H.; Dragoi, I.; Sellers  
 L., R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.  
 Nature 406, 477-483, 2000  
 A:Title: DNA Sequence of both chromosomes of the cholera pathogen *Vibrio cholerae*.  
 #:Reference number: A82035; MUID:2040683; PMID:10952301

A:Residues: 1-182 <HRF>  
A:Cross-references: GB:AE004195; GB:AE003852; NID:g9655621; PIDN:AAF94313.1; GSPDB:G  
A:Experimental source: serogroup O1; strain N16561; biotype El Tor  
C:Genetics:  
A:Gene: VC1154  
A:Map position: 1  
C:Superfamily: Vibrio cholerae hypothetical protein VC1154

Query Match                53.0%; Score 44; DB 2; Length 182;  
Best Local Similarity     57.9%; Pred. No. 19;  
Matches    11; Conservative    1; Mismatches        7; Indels    0; Caps      0;

OY            1 MRATSLAAALAVAGDALA 19  
||| | | | |||||  
Db          1 MKRTIVATALLVAGSALA 19

```

RESULT 12
AI3234
hypothetical protein ssua (imported) - Agrobacterium tumefaciens (strain C58, Dupont)
C:Species: Agrobacterium tumefaciens
C:Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 11-Jan-2002
C:Accession: AI3234

```

Matches 11: Conservative 2; Mismatches 6; Indels 0; Gaps 0;

QY 1 MRATSLAAALAVAGDALA 19  
 :|||||:|||||  
 Db 5 LRMVTLAAAVGVAGNMLA 23

RESULT 3  
 E64002  
 hypothetical protein H10131 - Haemophilus influenzae (strain Rd KW20)

C:Species: Haemophilus influenzae  
 C>Date: 18-Aug-1995 #sequence\_revision 18-Aug-1995 #text\_change 21-Jul-2000  
 C:Accession: E64002  
 R:Feilschmann, R.D.; Adams, M.D.; White, O.; Clayton, R.A.; Kirkness, E.F.; Kerlavage, A.; Gocayne, J.D.; Scott, J.; Shirley, R.; Liu, L.I.; Goddek, A.; Kelley, J.M.; Weidman, D.M.; Brandon, R.C.; Fine, L.D.; Fritchman, J.L.; Fuhrmann, J.L.; Geoghegan, N.S.M. Science 269, 496-512, 1995  
 A:Authors: Gnehm, C.L.; McDonald, L.A.; Small, K.V.; Fraser, C.M.; Smith, H.O.; Venter, A.; Title: Whole-genome random sequencing and assembly of Haemophilus influenzae Rd.  
 A:Reference number: A64000; MUID:95350630; PMID:7542800  
 A:Accession: E64002  
 A:Status: nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-346 <TIGR>  
 A:Cross-references: GB:U032699; GB:U42023; NID:93212180; PIDN:AAC21803.1; PID:91573085;

Query Match  
 Best Local Similarity 56.6%; Score 47; DB 2; Length 346;  
 Matches 11: Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 3 ATSLAALAVAGDALA 19  
 :|||||:|||||  
 Db 11 STALLAAGLVGSGANA 27

## RESULT 4

P95868  
 conserved hypothetical protein [Imported] - Sinorhizobium meliloti (strain 1021), magapla  
 C:Species: Sinorhizobium meliloti  
 C>Date: 24-Aug-2001 #sequence\_revision 24-Aug-2001 #text\_change 30-Sep-2001  
 C:Accession: P95868  
 R:Finan, T.M.; Weidner, S.; Wong, K.; Buhrmester, J.; Chain, P.; Vorholter, F.J.; Hernat  
 Proc. Natl. Acad. Sci. U.S.A. 98, 9869-9894, 2001  
 A:Title: The complete sequence of the 1,683-kb PSYMB megaplasmid from the N2-fixing endo  
 A:Reference number: A95842; MUID:21396508; PMID:11481431  
 A:Accession: P95868  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-136 <KUR>

A:Cross-references: GB:AL591965; PIDN:CAC48614.1; PID:915140086; GSPDB:GN00167  
 A:Experimental source: strain 1021, megaplasmid PSYMB  
 R:Gilbert, F.; Finan, T.M.; Long, S.R.; Puhler, A.; Abola, P.; Ampe, F.; Barloy-Hubler, P.; Chain, P.; Cowie, A.; Davis, R.W.; Dreano, S.; Federspiel, N.A.; Fisher, R.F.; L.; Hyman, R.W.; Jones, T. Science 293, 668-672, 2001  
 A:Authors: Kahn, D.; Kahn, M.L.; Kalman, S.; Keating, D.H.; Kiss, E.; Komp, C.; LeJaure, hebullet, P.; Vandenbol, M.; Vorholter, F.J.; Weidner, S.; Wells, D.H.; Wong, K.; Yeh, K. A:Title: The composite genome of the legume symbiont Sinorhizobium meliloti.  
 A:Reference number: A96039; MUID:21368234; PMID:11474104  
 A:Contents: annotation  
 C:Genetics:  
 A:Gene: SMD20222  
 A:Genome: plasmid

Query Match  
 Best Local Similarity 55.4%; Score 46; DB 2; Length 136;  
 Matches 10: Conservative 3; Mismatches 6; Indels 0; Gaps 0;

QY 1 MRATSLAALAVAGDALA 19  
 :|||||:|||||  
 Db 4 MKTSRLAALAVVGSGLA 22

RESULT 5  
 T31888  
 hypothetical protein C03A7.7 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans  
 C>Date: 29-Oct-1999 #sequence\_revision 29-Oct-1999 #text\_change 21-Jan-2000  
 C:Accession: T31888  
 R:Greco, T.; Bradshaw, H.; Elliott, G. submitted to the EMBL Data Library, July 1997  
 A:Description: The sequence of C. elegans cosmid C03A7.  
 A:Reference number: Z21096  
 A:Accession: T31888  
 A:Status: preliminary; translated from GB/EMBL/DDBJ  
 A:Molecule type: DNA  
 A:Residues: 1-388 <GRE>  
 A:Cross-references: EMBL:AF016451; PIDN:AA85996.1; GSPDB:GN00023; CESP:C03A7.7  
 A:Experimental source: strain Bristol N2, clone C03A7  
 C:Genetics:  
 A:Gene: CESP:C03A7.7  
 A:Map position: 5  
 A:Introns: 75/3  
 C:Superfamily: gliadin

Query Match  
 Best Local Similarity 55.4%; Score 46; DB 2; Length 388;  
 Matches 13: Conservative 2; Mismatches 3; Indels 4; Gaps 1;

QY 1 MRATSLAALAVAGDALA 18  
 :|||||:|||||  
 Db 1 MRFSLAIFLACALAVSGSAI 22

## RESULT 6

D96551  
 hypothetical protein F1IM15.20 [Imported] - Arabidopsis thaliana  
 C:Species: Arabidopsis thaliana (mouse-ear cress)  
 C>Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 31-Mar-2001  
 C:Accession: D96551  
 R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alon  
 Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Creasy, T.H.; Dewar, ansen, N.F.; Hughes, B.; Hultzer, L. Nature 408, 816-820, 2000  
 A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Matli, R.; Marzia  
 Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.  
 A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shim, P.; Southwick, A.M.; Sun, H.; Tallo  
 ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.  
 A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.  
 A:Reference number: A86141; MUID:21016719; PMID:11130712  
 A:Accession: D96551  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-501 <STO>  
 A:Cross-references: GB:AE005173; NID:94836944; PIDN:AAD30646.1; GSPDB:GN00141  
 C:Genetics:  
 A:Gene: F1IM15.20  
 A:Map position: 1

Query Match  
 Best Local Similarity 55.4%; Score 46; DB 2; Length 501;  
 Matches 11: Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 3 ATSLAALAVAGDALA 18  
 :|||||:|||||  
 Db 352 ATSLADGVAVAGQAI 367

## RESULT 7

F86767  
 hypothetical protein ylfF [Imported] - Lactococcus lactis subsp. lactis (strain IL140  
 C:Species: Lactococcus lactis subsp. lactis  
 C>Date: 23-Mar-2001 #sequence\_revision 23-Mar-2001 #text\_change 03-Aug-2001  
 C:Accession: F86767  
 R:Boletcin, A.; Wincker, P.; Manger, S.; Jallion, O.; Malarme, K.; Weissenbach, J.; Eh

GenCore version 5.1.4-p5\_4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 9, 2003, 15:16:46 ; Search time 1.25799 Seconds  
(without alignments)  
1451.961 Million cell updates/sec

Title: US-10-028-245-3  
Perfect score: 83  
Sequence: 1 MRATSLAALAVAGDALA 19

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :  
1: p1r1:\*  
2: p1r2:\*  
3: p1r3:\*  
4: p1r4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	49	59.0	183	2 B87452	hypothetical prote
2	47	56.6	202	2 A83448	hypothetical prote
3	47	56.6	346	2 E64002	hypothetical prote
4	46	55.4	136	2 F95868	conserved hypotet
5	46	55.4	388	2 T31888	hypothetical prote
6	45	54.2	501	2 D96551	hypothetical prote
7	45	54.2	166	2 F86767	hypothetical prote
8	45	54.2	178	2 E87300	conserved hypotet
9	45	54.2	363	2 T34931	hypothetical prote
10	45	54.2	472	2 H70836	hypothetical prote
11	44	53.0	182	2 E82235	hypothetical prote
12	44	53.0	320	2 A13234	hypothetical prote
13	43	51.8	89	2 A11033	hypothetical prote
14	43	51.8	100	1 GEH1	hypothetical prote
15	43	51.8	137	2 C84543	probable Hesb-like
16	43	51.8	213	2 C96016	conserved hypotet
17	43	51.8	243	2 G75624	hypothetical prote
18	43	51.8	245	2 A12126	hypothetical prote
19	43	51.8	272	2 I52690	probabilin - human
20	43	51.8	272	2 A39682	probabilin - rat
21	43	51.8	292	2 AH0998	heat shock protein
22	43	51.8	368	2 F75546	conserved hypotet
23	43	51.8	498	2 A75509	ABC transporter, A
24	42	50.6	80	2 T09045	hypothetical prote
25	42	50.6	306	2 AC0602	probable exported
26	42	50.6	332	2 A42171	A2-adenosine recep
27	42	50.6	332	2 I48933	adenosine receptor
28	42	50.6	388	2 T31887	hypothetical prote
29	42	50.6	429	2 D70763	threonine ammonia-

30	42	50.6	513	2 T34546	hypothetical prote
31	42	50.6	888	2 T31131	pyruvate, phosphat
32	42	50.6	1125	2 H87644	TonB-dependent rec
33	41.5	50.0	128	1 A57321	E48 antigen precur
34	41	49.4	268	2 AE2458	hypothetical prote
35	41	49.4	346	2 E83121	D-alanine-D-alanin
36	41	49.4	348	2 AC2279	hypothetical prote
37	41	49.4	349	2 H87295	A/G-specific adenyl
38	41	49.4	380	2 S14188	carbonate dehydrat
39	41	49.4	501	2 AE3614	succinoglycan bios
40	41	49.4	1345	2 T00964	hypothetical prote
41	41	49.4	1819	2 D97132	uncharacterized ph
42	40.5	48.8	203	2 D81934	probable periplasm
43	40.5	48.8	203	2 F81171	cryptic protein NM
44	40	48.2	79	2 C87490	hypothetical prote
45	40	48.2	91	2 T08131	oleosin-like prote

## ALIGNMENTS

```

RESULT 1
B87452
hypothetical protein CC1636 [imported] - Caulobacter crescentus
C:Species: Caulobacter crescentus
C:Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 20-Apr-2001
C:Accession: B87452
R.Nierman, W.C.; Feldblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg,
B.; Laub, M.T.; Deboy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Ko
n, J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C
Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001
A:Title: Complete Genome Sequence of Caulobacter crescentus.
A:Reference number: A87249; MUID:21173698; PMID:11259647
A:Accession: B87452
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-183 <STO>
A:Cross-references: GB:AE005673; NID:g13423040; PIDN:AK23614.1; GSPDB:GN00148
C:Genetics:
A:Gene: CC1636

Query Match          59.0%; Score 49; DB 2; Length 183;
Best Local Similarity 57.9%; Pred. No. 3.3;
Matches 11; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

QY      1 MRATSLAALAVAGDALA 19
Db      37 MEVASTIAAGLSVAGSALA 55

RESULT 2
A83448
hypothetical protein PA1579 [imported] - Pseudomonas aeruginosa (strain PA01)
C:Species: Pseudomonas aeruginosa
C:Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
C:Accession: A83448
R.Stover, C.K.; Pham, X.Q.; EWlin, A.L.; Mizoguchi, S.D.; Warriner, P.; Hickey, M.J.; L
adman, S.; Vuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larity, K.; L
oy, S.; Olson, M.V.
Nature 406, 959-964, 2000
A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic pa
A:Reference number: A82950; MUID:20437337; PMID:10984043
A:Accession: A83448
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-202 <STO>
A:Cross-references: GB:AE004586; GB:AE004091; NID:g9947536; PIDN:AA04968.1; GSPDB:GN
C:Genetics:
A:Gene: PA1579

Query Match          56.6%; Score 47; DB 2; Length 202;
Best Local Similarity 57.9%; Pred. No. 7.3;

```

NUMBER OF SEQ ID NOS: 292  
 SOFTWARE: fastseq for Windows Version 4.0  
 SEQ ID NO 259  
 LENGTH: 1289  
 TYPE: PRT  
 ORGANISM: Mycobacterium tuberculosis  
 US-09-712-363-259

Query Match 48.2%; Score 40; DB 9; Length 1289;  
 Best Local Similarity 76.9%; Pred. No. 5.8e+02;  
 Matches 10; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
 QY 7 LAALAVAGDALA 19  
 Db. 617 LAALGPADALA 629

Search completed: May 9, 2003, 15:17:21  
 Job time : 2.47489 secs

```

; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 11983
; LENGTH: 346
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-815-242-11983

Query Match          49.4%; Score 41; DB 10; Length 346;
Best Local Similarity 57.9%; Pred. No. 1e+02;
Matches 11; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

OY 1 MRATSLAALAVAGDALA 19
Db 306 MRATSLPOLAAGIGFA 324

RESULT 12
US-09-867-550-306
; Sequence 306, Application US/09867550
; Patent No. US20020082206A1
; GENERAL INFORMATION:
; APPLICANT: Leach, Martin D.
; APPLICANT: Mehraban, Fuad,
; APPLICANT: Conley, Pamela
; APPLICANT: Law, Debbie
; APPLICANT: Topper, James
; TITLE OF INVENTION: No. US20020082206A1 Polynucleotides from Atherogenic Cells and
; FILE REFERENCE: 21402-013 (Cura-313)
; CURRENT APPLICATION NUMBER: US/09/867,550
; CURRENT FILING DATE: 2001-09-20
; PRIOR APPLICATION NUMBER: USSN 60/208,427
; PRIOR FILING DATE: 2000-05-30
; NUMBER OF SEQ ID NOS: 2125
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 306
; LENGTH: 73
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-867-550-306

Query Match          48.2%; Score 40; DB 10; Length 73;
Best Local Similarity 60.0%; Pred. No. 27;
Matches 9; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

OY 1 MRATSLAALAVAG 15
Db 1 MKRTSLIIVLALAG 15

RESULT 13
US-10-125-540-292
; Sequence 292, Application US/10125540
; Publication No. US20030059875A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PT214C1
; CURRENT APPLICATION NUMBER: US/10/125,540
; CURRENT FILING DATE: 2002-04-19
; Prior Application removed - See file wrapper or Palm
; NUMBER OF SEQ ID NOS: 646
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 292
```

```

; LENGTH: 293
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-125-540-292

Query Match          48.2%; Score 40; DB 9; Length 293;
Best Local Similarity 81.8%; Pred. No. 1.2e+02;
Matches 9; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 6 LLAALAVAGD 16
Db 220 LLAALVVGVD 230

RESULT 14
US-09-764-870-292
; Sequence 292, Application US/09764870
; Patent No. US20020042386A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PT214
; CURRENT APPLICATION NUMBER: US/09/764,870
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 646
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 292
; LENGTH: 293
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-764-870-292

Query Match          48.2%; Score 40; DB 10; Length 293;
Best Local Similarity 81.8%; Pred. No. 1.2e+02;
Matches 9; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 6 LLAALAVAGD 16
Db 220 LLAALVVGVD 230

RESULT 15
US-09-712-363-259
; Sequence 259, Application US/09712363
; Patent No. US20020164588A1
; GENERAL INFORMATION:
; APPLICANT: Eisenberg, David
; APPLICANT: Rotstein, Sergio H.
; APPLICANT: Marcotte, Edward M.
; TITLE OF INVENTION: DETERMINING THE FUNCTIONS AND
; FILE REFERENCE: 07419-032001
; CURRENT APPLICATION NUMBER: US/09/712,363
; CURRENT FILING DATE: 2000-11-13
; PRIOR APPLICATION NUMBER: PCT/US00/02246
; PRIOR FILING DATE: 2000-01-28
; PRIOR APPLICATION NUMBER: 60/179,531
; PRIOR FILING DATE: 2000-02-01
; PRIOR APPLICATION NUMBER: 60/117,844
; PRIOR FILING DATE: 1999-01-29
; PRIOR APPLICATION NUMBER: 60/118,206
; PRIOR FILING DATE: 1999-02-01
; PRIOR APPLICATION NUMBER: 60/126,593
; PRIOR FILING DATE: 1999-03-26
; PRIOR APPLICATION NUMBER: 60/134,093
; PRIOR FILING DATE: 1999-05-14
; PRIOR APPLICATION NUMBER: 60/134,092
; PRIOR FILING DATE: 1999-05-14
; PRIOR APPLICATION NUMBER: 60/165,124
; PRIOR FILING DATE: 1999-11-12
; PRIOR APPLICATION NUMBER: 60/165,086
; PRIOR FILING DATE: 1999-11-12
```

LENGTH: 25  
TYPE: PRT  
ORGANISM: Corynebacterium ammoniagenes  
US-10-112-488-2

Query Match 50.6%; Score 42; DB 9; Length 25;  
Best Local Similarity 57.9%; Pred. No. 4.3;  
Matches 11; Conservative 1; Mismatches 7; Indels 0; Gaps 0;

Qy 1 MRATSLAALAVAGDALA 19  
|: ||||| || ||  
Db 1 MKRMKSLAALTVAGAMLA 19

RESULT 9  
US-09-873-880-34  
Sequence 34, Application US/09873880  
Patent No. US20020123118A1  
GENERAL INFORMATION:  
APPLICANT: Sewalt, Vincent  
APPLICANT: Falco, S. Carl  
APPLICANT: Allen, Stephen M.  
TITLE OF INVENTION: GLYCINE METABOLISM ENZYMES  
FILE REFERENCE: BH1192 US CIP  
CURRENT APPLICATION NUMBER: US/09/873,880  
CURRENT FILING DATE: 2001-06-04  
PRIOR APPLICATION NUMBER: 09/363,321  
PRIOR FILING DATE: July 28, 1999  
PRIOR APPLICATION NUMBER: 60/094,839  
PRIOR FILING DATE: July 31, 1998  
NUMBER OF SEQ ID NOS: 42  
SOFTWARE: Microsoft Office 97  
SEQ ID NO 34  
LENGTH: 296  
TYPE: PRT  
ORGANISM: Zea mays  
US-09-873-880-34

Query Match 50.6%; Score 42; DB 10; Length 296;  
Best Local Similarity 55.6%; Pred. No. 60;  
Matches 10; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

Qy 2 RATSLAALAVAGDALA 19  
|: ||||| || ||  
Db 48 RSRSLAALAEISKDGS 65

RESULT 10  
US-09-864-761-43102  
Sequence 43102, Application US/09864761  
Patent No. US20020048763A1  
GENERAL INFORMATION:  
APPLICANT: Penn, Sharon G.  
APPLICANT: Rank, David R.  
APPLICANT: Hanzel, David K.  
APPLICANT: Chen, Wensheng  
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR  
FILE REFERENCE: Aeomica-X-1  
CURRENT APPLICATION NUMBER: US/09/864,761  
CURRENT FILING DATE: 2001-05-23  
PRIOR APPLICATION NUMBER: US 60/180,312  
PRIOR FILING DATE: 2000-02-04  
PRIOR APPLICATION NUMBER: US 60/207,456  
PRIOR FILING DATE: 2000-05-26  
PRIOR APPLICATION NUMBER: US 09/632,366  
PRIOR FILING DATE: 2000-08-03  
PRIOR APPLICATION NUMBER: GB 24263.6  
PRIOR FILING DATE: 2000-10-04  
PRIOR APPLICATION NUMBER: US 60/236,359  
PRIOR FILING DATE: 2000-09-27  
PRIOR APPLICATION NUMBER: PCT/US01/00666  
PRIOR FILING DATE: 2001-01-30

PRIOR APPLICATION NUMBER: PCT/US01/00667  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00664  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00669  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00665  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00668  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00663  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00662  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00661  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00670  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: US 60/234,687  
PRIOR FILING DATE: 2000-09-21  
PRIOR APPLICATION NUMBER: US 09/608,408  
PRIOR FILING DATE: 2000-06-30  
PRIOR APPLICATION NUMBER: US 09/774,203  
PRIOR FILING DATE: 2001-01-29  
NUMBER OF SEQ ID NOS: 49117  
SOFTWARE: Annomax Sequence Listing Engine vers. 1.1  
SEQ ID NO 43102  
LENGTH: 712  
TYPE: PRT  
ORGANISM: Homo sapiens  
FEATURE:  
OTHER INFORMATION: MAP TO AC005880.2  
OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.4  
OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 0.95  
OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1  
OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 1.2  
OTHER INFORMATION: EST\_HUMAN HIT: AW389729.1, EVALUE 1.00e-112  
OTHER INFORMATION: SWISSPROT HIT: P05130, EVALUE 9.00e-06  
US-09-864-761-43102

Query Match 50.6%; Score 42; DB 10; Length 712;  
Best Local Similarity 58.8%; Pred. No. 1.5e+02;  
Matches 10; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

Qy 2 RATSLAALAVAGDAL 18  
|: ||||| || ||  
Db 633 RKSLAALAKSGERL 649

RESULT 11  
US-09-815-242-11983  
Sequence 11983, Application US/09815242  
Patent No. US20020061569A1  
GENERAL INFORMATION:  
APPLICANT: Haselbeck, Robert  
APPLICANT: Ohlsen, Karl L.  
APPLICANT: Zyskind, Judith W.  
APPLICANT: Wall, Daniel  
APPLICANT: Trawick, John D.  
APPLICANT: Carr, Grant J.  
APPLICANT: Yamamoto, Robert T.  
APPLICANT: Xu, H. Howard  
TITLE OF INVENTION: Identification of Essential Genes in  
FILE REFERENCE: ELITRA.011A  
CURRENT APPLICATION NUMBER: US/09/815,242  
CURRENT FILING DATE: 2001-03-21  
PRIOR APPLICATION NUMBER: 60/191,078  
PRIOR FILING DATE: 2000-03-21  
PRIOR APPLICATION NUMBER: 60/206,848  
PRIOR FILING DATE: 2000-05-23  
PRIOR APPLICATION NUMBER: 60/207,727  
PRIOR FILING DATE: 2000-05-26

Matches 10: Conservative 0: Mismatches 4: Indels 0: Gaps 0:  
QY 3 ATSLAALAVAGD 16  
|||||  
Db 53 ATSLTATLTKFAGD 66

## RESULT 5

US-09-925-301-1319  
; Sequence 1319, Application US/09925301  
; Patent No. US20020052308A1  
; GENERAL INFORMATION:  
; APPLICANT: Rosen et al.  
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies  
; FILE REFERENCE: PA106  
; CURRENT APPLICATION NUMBER: US/09/925,301  
; PRIOR FILING DATE: 2001-08-10  
; PRIOR APPLICATION NUMBER: PCT/US00/05882  
; PRIOR FILING DATE: 2000-03-08  
; PRIOR APPLICATION NUMBER: 60/124,270  
; PRIOR FILING DATE: 1999-03-12  
; NUMBER OF SEQ ID NOS: 1694  
; SOFTWARE: Patentln Ver. 2.0  
; SEQ ID NO 1319  
; LENGTH: 279  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-925-301-1319

Query Match 51.8%; Score 43; DB 10; Length 279;  
Best Local Similarity 52.9%; Pred. No. 40;  
Matches 9; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

QY 2 RATSILAALAVAGD 18  
:|:|:|:|:|:|  
Db 226 KAAELIANSLATAGDGL 242

## RESULT 6

US-09-815-242-13808  
; Sequence 13808, Application US/09815242  
; Patent No. US20020061569A1  
; GENERAL INFORMATION:  
; APPLICANT: Haselbeck, Robert  
; APPLICANT: Ohlsen, Karl L.  
; APPLICANT: Zyskind, Judith W.  
; APPLICANT: Wall, Daniel  
; APPLICANT: Trawick, John D.  
; APPLICANT: Carr, Grant J.  
; APPLICANT: Yamamoto, Robert T.  
; APPLICANT: Xu, H. Howard  
; TITLE OF INVENTION: Identification of Essential Genes in  
; FILE REFERENCE: ELITRA.011A  
; CURRENT APPLICATION NUMBER: US/09/815,242  
; PRIOR FILING DATE: 2001-03-21  
; PRIOR APPLICATION NUMBER: 60/191,078  
; PRIOR FILING DATE: 2000-03-21  
; PRIOR APPLICATION NUMBER: 60/206,848  
; PRIOR FILING DATE: 2000-05-23  
; PRIOR APPLICATION NUMBER: 60/207,727  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: 60/242,578  
; PRIOR FILING DATE: 2000-10-23  
; PRIOR APPLICATION NUMBER: 60/253,625  
; PRIOR FILING DATE: 2000-11-27  
; PRIOR APPLICATION NUMBER: 60/257,931  
; PRIOR FILING DATE: 2000-12-22  
; PRIOR APPLICATION NUMBER: 60/269,308  
; PRIOR FILING DATE: 2001-02-16  
; NUMBER OF SEQ ID NOS: 14110  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 13808

; LENGTH: 294  
; TYPE: PRT  
; ORGANISM: Salmonella typhi  
US-09-815-242-13808

Query Match 51.8%; Score 43; DB 10; Length 294;  
Best Local Similarity 71.4%; Pred. No. 42;  
Matches 10; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 3 ATSLAALAVAGD 16  
|||||  
Db 53 ATSLTATLTKFAGD 66

## RESULT 7

US-09-935-642-4  
; Sequence 4, Application US/09935642  
; Publication No. US20030044795A1  
; GENERAL INFORMATION:  
; APPLICANT: BYRJALSEN, Inger  
; APPLICANT: LARSEN, Peter  
; APPLICANT: STEPHEN, John  
; TITLE OF INVENTION: Biochemical Markers for the Human  
; FILE REFERENCE: 8969-014  
; CURRENT APPLICATION NUMBER: US/09/935,642  
; PRIOR FILING DATE: 2001-08-24  
; PRIOR APPLICATION NUMBER: PCT/GB97/02394  
; PRIOR FILING DATE: 1997-09-05  
; PRIOR APPLICATION NUMBER: PCT/GB97/07132.8  
; PRIOR FILING DATE: 1997-04-08  
; PRIOR APPLICATION NUMBER: PCT/GB9618600.2  
; PRIOR FILING DATE: 1996-09-06  
; NUMBER OF SEQ ID NOS: 16  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 4  
; LENGTH: 471  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-935-642-4

Query Match 51.8%; Score 43; DB 9; Length 471;  
Best Local Similarity 52.9%; Pred. No. 70;  
Matches 9; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

QY 2 RATSILAALAVAGD 18  
:|:|:|:|:|:|  
Db 418 KAAELIANSLATAGDGL 434

## RESULT 8

US-10-112-488-2  
; Sequence 2, Application US/10112488  
; Publication No. US20030082746A1  
; GENERAL INFORMATION:  
; APPLICANT: KIKUCHI, Yoshimi  
; APPLICANT: DATE, Masayo  
; APPLICANT: UMEZAWA, Yukiko  
; APPLICANT: YOKOTAMA, Keiichi  
; APPLICANT: MATSUI, Hiroshi  
; TITLE OF INVENTION: PROCESS FOR PRODUCING TRANSGLUTAMINASE  
; FILE REFERENCE: 219286USOCNT  
; CURRENT APPLICATION NUMBER: US/10/112,488  
; PRIOR FILING DATE: 2002-04-01  
; PRIOR APPLICATION NUMBER: PCT/JP00/06780  
; PRIOR FILING DATE: 2000-09-29  
; PRIOR APPLICATION NUMBER: JP2000-280098  
; PRIOR FILING DATE: 2000-06-28  
; PRIOR APPLICATION NUMBER: JP11-280098  
; PRIOR FILING DATE: 1999-09-30  
; NUMBER OF SEQ ID NOS: 70  
; SOFTWARE: Patentln version 3.1  
; SEQ ID NO 2

APPLICANT: SMITHSON, GLENNDA  
TITLE OF INVENTION: NOVEL HUMAN PROTEINS, POLYNUCLEOTIDES ENCODING THEM AND  
FILE REFERENCE: 15966-776  
CURRENT APPLICATION NUMBER: US/09/839,446  
PRIOR FILING DATE: 2001-04-19  
PRIOR APPLICATION NUMBER: 60/198,293  
PRIOR FILING DATE: 2000-04-19  
PRIOR APPLICATION NUMBER: 60/198,645  
PRIOR FILING DATE: 2000-04-20  
PRIOR APPLICATION NUMBER: 60/210,809  
PRIOR FILING DATE: 2000-06-09  
PRIOR APPLICATION NUMBER: 60/199,476  
PRIOR FILING DATE: 2000-04-26  
PRIOR APPLICATION NUMBER: 60/200,025  
PRIOR FILING DATE: 2000-04-26  
PRIOR APPLICATION NUMBER: 60/224,610  
PRIOR FILING DATE: 2000-08-11  
PRIOR APPLICATION NUMBER: 60/200,024  
PRIOR FILING DATE: 2000-04-26  
PRIOR APPLICATION NUMBER: 60/199,880  
PRIOR FILING DATE: 2000-04-26  
PRIOR APPLICATION NUMBER: 60/218,591  
PRIOR FILING DATE: 2000-07-17  
PRIOR APPLICATION NUMBER: 60/271,814  
PRIOR FILING DATE: 2001-02-27  
NUMBER OF SEQ ID NOS: 49  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 34  
LENGTH: 271  
TYPE: PRT  
ORGANISM: Unknown Organism  
FEATURE:  
OTHER INFORMATION: Description of Unknown Organism: POLYX  
OTHER INFORMATION: GM\_11817402\_A  
US-09-839-446-34

Query Match 53.08; Score 44; DB 9; Length 271;  
Best Local Similarity 47.18; Pred. No. 28;  
Matches 8; Conservative 5; Mismatches 4; Indels 0; Gaps 0;

2 RATSLAALAVAGDAL 18  
Db 214 KATEFIASSVATAGDGL 230

RESULT 3  
US-09-898-570-34  
Sequence 34, Application US/09898570  
Patent No. US20020123612A1  
GENERAL INFORMATION:  
APPLICANT: GERLACH, VALERIE L.  
APPLICANT: ELLERMAN, KAREN  
APPLICANT: MACDOUGALL, JOHN R.  
APPLICANT: SMITHSON, GLENNDA  
TITLE OF INVENTION: NOVEL HUMAN PROTEINS, POLYNUCLEOTIDES ENCODING THEM AND  
FILE REFERENCE: 15966-776CIP  
CURRENT APPLICATION NUMBER: US/09/898,570  
PRIOR FILING DATE: 2001-07-03  
PRIOR APPLICATION NUMBER: 60/198,293  
PRIOR FILING DATE: 2000-04-19  
PRIOR APPLICATION NUMBER: 60/198,645  
PRIOR FILING DATE: 2000-04-20  
PRIOR APPLICATION NUMBER: 60/210,809  
PRIOR FILING DATE: 2000-06-09  
PRIOR APPLICATION NUMBER: 60/199,476  
PRIOR FILING DATE: 2000-04-26  
PRIOR APPLICATION NUMBER: 60/200,025  
PRIOR FILING DATE: 2000-04-26  
PRIOR APPLICATION NUMBER: 60/224,610  
PRIOR FILING DATE: 2000-08-11  
PRIOR APPLICATION NUMBER: 60/200,024

PRIOR FILING DATE: 2000-04-26  
PRIOR APPLICATION NUMBER: 60/199,880  
PRIOR FILING DATE: 2000-04-26  
PRIOR APPLICATION NUMBER: 60/218,591  
PRIOR FILING DATE: 2000-07-17  
PRIOR APPLICATION NUMBER: 60/271,814  
PRIOR FILING DATE: 2001-02-27  
PRIOR APPLICATION NUMBER: 60/215,855  
PRIOR FILING DATE: 2000-07-03  
PRIOR APPLICATION NUMBER: 09/839,446  
PRIOR FILING DATE: 2001-04-19  
NUMBER OF SEQ ID NOS: 58  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 34  
LENGTH: 271  
TYPE: PRT  
ORGANISM: Unknown Organism  
FEATURE:  
OTHER INFORMATION: Description of Unknown Organism: POLYX  
OTHER INFORMATION: GM\_11817402\_A  
US-09-898-570-34

Query Match 53.08; Score 44; DB 10; Length 271;  
Best Local Similarity 47.18; Pred. No. 28;  
Matches 8; Conservative 5; Mismatches 4; Indels 0; Gaps 0;

2 RATSLAALAVAGDAL 18  
Db 214 KATEFIASSVATAGDGL 230

RESULT 4  
US-09-815-242-11709  
Sequence 11709, Application US/09815242  
Patent No. US20020061569A1  
GENERAL INFORMATION:  
APPLICANT: Haselbeck, Robert  
APPLICANT: Ohlsen, Karl L.  
APPLICANT: Zyskind, Judith W.  
APPLICANT: Wall, Daniel  
APPLICANT: Trawick, John D.  
APPLICANT: Carr, Grant J.  
APPLICANT: Yamamoto, Robert T.  
APPLICANT: Xu, H. Howard  
TITLE OF INVENTION: Identification of Essential Genes in  
FILE REFERENCE: ELITRA 011A  
CURRENT APPLICATION NUMBER: US/09/815,242  
PRIOR FILING DATE: 2001-03-21  
PRIOR APPLICATION NUMBER: 60/191,078  
PRIOR FILING DATE: 2000-03-21  
PRIOR APPLICATION NUMBER: 60/206,848  
PRIOR FILING DATE: 2000-05-23  
PRIOR APPLICATION NUMBER: 60/207,727  
PRIOR FILING DATE: 2000-05-26  
PRIOR APPLICATION NUMBER: 60/242,578  
PRIOR FILING DATE: 2000-10-23  
PRIOR APPLICATION NUMBER: 60/253,625  
PRIOR FILING DATE: 2000-11-27  
PRIOR APPLICATION NUMBER: 60/257,931  
PRIOR FILING DATE: 2000-12-22  
PRIOR APPLICATION NUMBER: 60/269,308  
PRIOR FILING DATE: 2001-02-16  
NUMBER OF SEQ ID NOS: 1410  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 11709  
LENGTH: 226  
TYPE: PRT  
ORGANISM: Klebsiella pneumoniae  
US-09-815-242-11709

Query Match 51.88; Score 43; DB 10; Length 226;  
Best Local Similarity 71.48; Pred. No. 32;

GenCore version 5.1.4-PS.4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 9, 2003, 15:09:30 ; Search time 1.47489 Seconds  
(without alignments)  
1185.508 Million cell updates/sec

Title: US-10-028-245-3

Perfect score: 83

Sequence: 1 MRATSLAALAVAGDALA 19

Scoring table:

BLOSUM62

Gapop 10.0, Gapext 0.5

Searched: 349150 seqs, 92025710 residues

Total number of hits satisfying chosen parameters: 349150

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 10%

Listing first 45 summaries

Database: Published Applications\_A1:

1: /cgn2\_6/ptodata/2/pubppa/US08\_NEW\_PUB.pep:\*  
2: /cgn2\_6/ptodata/2/pubppa/PCT\_NEW\_PUB.pep:\*  
3: /cgn2\_6/ptodata/2/pubppa/US06\_NEW\_PUB.pep:\*  
4: /cgn2\_6/ptodata/2/pubppa/US06\_PUBCOMB.pep:\*  
5: /cgn2\_6/ptodata/2/pubppa/US07\_NEW\_PUB.pep:\*  
6: /cgn2\_6/ptodata/2/pubppa/US07\_PUBCOMB.pep:\*  
7: /cgn2\_6/ptodata/2/pubppa/PCTUS\_PUBCOMB.pep:\*  
8: /cgn2\_6/ptodata/2/pubppa/US08\_PUBCOMB.pep:\*  
9: /cgn2\_6/ptodata/2/pubppa/US09\_PUBCOMB.pep:\*  
10: /cgn2\_6/ptodata/2/pubppa/US09\_PUBCOMB.pep:\*  
11: /cgn2\_6/ptodata/2/pubppa/US10\_PUBCOMB.pep:\*  
12: /cgn2\_6/ptodata/2/pubppa/US10\_PUBCOMB.pep:\*  
13: /cgn2\_6/ptodata/2/pubppa/US60\_NEW\_PUB.pep:\*  
14: /cgn2\_6/ptodata/2/pubppa/US60\_PUBCOMB.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	47	56.6	346	9 US-10-260-877-68	Sequence 68, Appl
2	47	53.0	271	9 US-09-839-446-34	Sequence 34, Appl
3	44	53.0	271	10 US-09-898-570-34	Sequence 11709, A
4	43	51.8	226	10 US-09-815-242-11709	Sequence 1319, A
5	43	51.8	279	10 US-09-925-301-1319	Sequence 13808, A
6	43	51.8	294	10 US-09-815-242-13808	Sequence 4, Appl
7	43	51.8	471	9 US-09-935-642-4	Sequence 11963, A
8	42	50.6	25	9 US-10-112-488-2	Sequence 34, Appl
9	42	50.6	296	10 US-09-873-880-34	Sequence 43102, A
10	42	50.6	712	10 US-09-864-761-43102	Sequence 11963, A
11	41	49.4	346	10 US-09-815-242-11983	Sequence 306, Appl
12	40	48.2	73	10 US-09-867-550-306	Sequence 292, Appl
13	40	48.2	293	10 US-10-125-540-292	Sequence 259, Appl
14	40	48.2	293	10 US-09-764-870-292	Sequence 5698, Appl
15	40	48.2	1289	9 US-09-738-626-5698	Sequence 582, Appl
16	39.5	47.6	263	9 US-09-738-626-5698	Sequence 686, Appl
17	39	47.0	25	9 US-10-097-065-584	
18	39	47.0	139	9 US-10-097-065-582	
19	39	47.0	255	9 US-09-764-868-686	

20	39	47.0	299	10 US-09-815-242-10205	Sequence 10205, A
21	39	47.0	315	9 US-09-935-868-16	Sequence 16, Appl
22	39	47.0	332	10 US-09-815-242-5088	Sequence 5088, Ap
23	39	47.0	340	10 US-09-815-242-11980	Sequence 11980, A
24	39	47.0	360	9 US-09-935-868-15	Sequence 15, Appl
25	39	47.0	376	10 US-09-925-301-1399	Sequence 1399, Ap
26	39	47.0	395	9 US-09-738-626-6644	Sequence 6644, Ap
27	39	47.0	440	9 US-09-738-626-6376	Sequence 6376, Ap
28	39	47.0	468	9 US-10-247-463-12	Sequence 12, Appl
29	39	47.0	592	9 US-09-935-868-8	Sequence 8, Appl
30	39	47.0	699	9 US-10-008-355-8	Sequence 8, Appl
31	39	47.0	712	9 US-10-008-355-2	Sequence 2, Appl
32	39	47.0	892	9 US-09-738-626-5307	Sequence 5307, Ap
33	39	47.0	1158	9 US-09-935-868-26	Sequence 26, Appl
34	39	47.0	1168	9 US-09-935-868-24	Sequence 24, Appl
35	39	47.0	1832	9 US-10-014-717-4	Sequence 4, Appl
36	38	45.8	95	10 US-09-917-340-76	Sequence 76, Appl
37	38	45.8	223	9 US-09-931-457A-34	Sequence 34, Appl
38	38	45.8	226	10 US-09-791-171-52	Sequence 52, Appl
39	38	45.8	322	9 US-10-162-012-42	Sequence 42, Appl
40	38	45.8	381	9 US-09-931-457A-72	Sequence 72, Appl
41	38	45.8	470	9 US-09-931-457A-68	Sequence 68, Appl
42	38	45.8	527	9 US-09-738-626-5523	Sequence 5523, Ap
43	38	45.8	752	9 US-09-712-363-252	Sequence 252, Appl
44	37	44.6	46	10 US-09-864-761-44060	Sequence 44060, A
45	37	44.6	293	10 US-09-815-242-11126	Sequence 11126, A

## ALIGNMENTS

RESULT 1  
US-10-260-877-68  
; Sequence 68, Application US/10260877  
; Publication NO. US20030021813A1  
GENERAL INFORMATION:  
; APPLICANT: Abbott Laboratories  
; APPLICANT: Chovan, Linda E.  
; APPLICANT: Hessler, Paul E.  
; APPLICANT: Reich, Karl A.  
; TITLE OF INVENTION: ESSENTIAL BACTERIA GENES AND GENOME  
; TITLE OF INVENTION: SCANNING IN HAEMOPHILUS INFLUENZAE FOR THE IDENTIFICATION OF  
; FILE REFERENCE: 6565, US, P1  
; CURRENT FILING DATE: 2002-09-30  
; PRIOR APPLICATION NUMBER: US/10/260,877  
; PRIOR FILING DATE: 2000-08-25  
; NUMBER OF SEQ ID NOS: 137  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 68  
; LENGTH: 346  
; TYPE: PRT  
; ORGANISM: H. Influenzae  
US-10-260-877-68  
Query Match 56.6%; Score 47; DB 9; Length 346;  
Best Local Similarity 64.7%; Pred. No. 13;  
Matches 11; Conservative 3; Mismatches 3; Indels 0; Gaps 0;  
OY 3 ATSLAALAVAGDALA 19  
:|||||:|:  
Db 11 STALLAGLVSGSANA 27  
RESULT 2  
US-09-839-446-34  
; Sequence 34, Application US/09839446  
; Publication NO. US20030050322A1  
GENERAL INFORMATION:  
; APPLICANT: GERLACH, VALERIE L.  
; APPLICANT: ELLERMAN, KAREN  
; APPLICANT: MACDOUGALL, JOHN R.

```
SOFTWARE: Word 97
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/080,704A
FILING DATE: 18 May 1998
CLASSIFICATION: 530
ATTORNEY/AGENT INFORMATION:
NAME: Parr, Richard S.
REGISTRATION NUMBER: 32,586
REFERENCE/DOCKET NUMBER: 18699DB
TELECOMMUNICATION INFORMATION:
TELEPHONE: (732)594-4958
TELEFAX: (732)594-4720
INFORMATION FOR SEQ ID NO: 23:
SEQUENCE CHARACTERISTICS:
LENGTH: 332 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-080-704A-23

Query Match          47.0%; Score 39; DB 4; Length 332;
Best Local Similarity 56.7%; Pred. No. 72;
Matches 9; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 3 ATSLAALAVAGDAL 18
   | | | | | | | | |
Db 12 ALELYVALSVAGNVL 27

RESULT 12
5171840-7
; Patent No. 5171840
; APPLICANT: KISHIMOTO, TADAMITSU
; TITLE OF INVENTION: RECEPTOR PROTEIN FOR HUMAN B CELL
; STIMULATORY FACTOR-2
; NUMBER OF SEQUENCES: 11
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/298,694
; FILING DATE: 19-JAN-1989
; SEQ ID NO: 7;
; LENGTH: 344
5171840-7

Query Match          47.0%; Score 39; DB 6; Length 344;
Best Local Similarity 66.7%; Pred. No. 75;
Matches 10; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 5 SLTAAALAVAGDALA 19
   : | | | | | | | | |
Db 7 ALTAALLAAGAAALA 21

RESULT 13
5480796-7
; Patent No. 5480796
; APPLICANT: KISHIMOTO, TADAMITSU
; TITLE OF INVENTION: ANTIBODIES AGAINST THE RECEPTOR PROTEIN
; FOR HUMAN B CELL STIMULATORY FACTOR-2
; NUMBER OF SEQUENCES: 8
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/907,650
; FILING DATE: 02-JUL-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 298,694
; FILING DATE: 19-JAN-1989
; SEQ ID NO: 7;
; LENGTH: 344
5480796-7

Query Match          47.0%; Score 39; DB 6; Length 344;
Best Local Similarity 66.7%; Pred. No. 75;
Matches 10; Conservative 1; Mismatches 4; Indels 0; Gaps 0;
```

```
QY 5 SLTAAALAVAGDALA 19
   : | | | | | | | | |
Db 7 ALTAALLAAGAAALA 21

RESULT 14
5171840-5
; Patent No. 5171840
; APPLICANT: KISHIMOTO, TADAMITSU
; TITLE OF INVENTION: RECEPTOR PROTEIN FOR HUMAN B CELL
; STIMULATORY FACTOR-2
; NUMBER OF SEQUENCES: 11
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/298,694
; FILING DATE: 19-JAN-1989
; SEQ ID NO: 5;
; LENGTH: 386
5171840-5

Query Match          47.0%; Score 39; DB 6; Length 386;
Best Local Similarity 66.7%; Pred. No. 85;
Matches 10; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 5 SLTAAALAVAGDALA 19
   : | | | | | | | | |
Db 7 ALTAALLAAGAAALA 21

RESULT 15
5480796-5
; Patent No. 5480796
; APPLICANT: KISHIMOTO, TADAMITSU
; TITLE OF INVENTION: ANTIBODIES AGAINST THE RECEPTOR PROTEIN
; FOR HUMAN B CELL STIMULATORY FACTOR-2
; NUMBER OF SEQUENCES: 8
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/907,650
; FILING DATE: 02-JUL-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 298,694
; FILING DATE: 19-JAN-1989
; SEQ ID NO: 5;
; LENGTH: 386
5480796-5

Query Match          47.0%; Score 39; DB 6; Length 386;
Best Local Similarity 66.7%; Pred. No. 85;
Matches 10; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 5 SLTAAALAVAGDALA 19
   : | | | | | | | | |
Db 7 ALTAALLAAGAAALA 21

Search completed: May 9, 2003, 15:10:48
Job time : 3.17123 secs
```

**THIS PAGE BLANK (USPTO)**

STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-349-696-23

Query Match 47.0%; Score 39; DB 1; Length 332;  
Best Local Similarity 56.2%; Pred. No. 72;  
Matches 9; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

OY 3 ATSLAALAVAGDAL 18  
1 : |||: |||: |  
Db 12 ALELVIALSVAGNVL 27

RESULT 9  
US-08-233-009-23  
Sequence 23, Application US/08233009  
Patent No. 5646156  
GENERAL INFORMATION:

APPLICANT: Jacobson, Marlene A  
APPLICANT: Johnson, Robert G  
APPLICANT: Salvatore, Christopher A  
TITLE OF INVENTION: INHIBITION OF EOSINOPHIL  
TITLE OF INVENTION: ACTIVATION THROUGH A3 ADENOSINE RECEPTOR ANTAGONISM  
NUMBER OF SEQUENCES: 56  
CORRESPONDENCE ADDRESS:

ADDRESSEE: Merck & Co., Inc.  
STREET: P.O. Box 2000  
CITY: Rahway  
STATE: New Jersey  
COUNTRY: United States  
ZIP: 07065

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/233,009  
FILING DATE: 25-APR-1994  
CLASSIFICATION: 424  
ATTORNEY/AGENT INFORMATION:

NAME: Bencen, Gerard H.  
REGISTRATION NUMBER: 35,746  
REFERENCE/DOCKET NUMBER: 19219  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (908) 594-4720  
TELEFAX: (908) 594-4720  
INFORMATION FOR SEQ ID NO: 23:  
SEQUENCE CHARACTERISTICS:

LENGTH: 332 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
FRAGMENT TYPE: N-terminal  
FEATURE:

NAME/KEY: Modified-site  
LOCATION: 216  
OTHER INFORMATION: /label= threonine  
US-08-233-009-23

Query Match 47.0%; Score 39; DB 1; Length 332;  
Best Local Similarity 56.2%; Pred. No. 72;  
Matches 9; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

OY 3 ATSLAALAVAGDAL 18  
1 : |||: |||: |  
Db 12 ALELVIALSVAGNVL 27

RESULT 10

US-08-560-231-23  
Sequence 23, Application US/08560231  
Patent No. 5817760  
GENERAL INFORMATION:

APPLICANT: Jacobson, Marlene A  
APPLICANT: Johnson, Robert G  
APPLICANT: Luneau, Christopher J  
APPLICANT: Salvatore, Christopher A  
TITLE OF INVENTION: Human Adenosine Receptors  
NUMBER OF SEQUENCES: 28  
CORRESPONDENCE ADDRESS:

ADDRESSEE: Merck & Co., Inc.  
STREET: P.O. Box 2000  
CITY: Rahway  
STATE: NJ  
COUNTRY: United States  
ZIP: 07065

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: Macintosh IIx1  
OPERATING SYSTEM: Macintosh  
SOFTWARE: Microsoft Word 5.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/560,231  
FILING DATE:

CLASSIFICATION: 530  
ATTORNEY/AGENT INFORMATION:  
NAME: Meredith, Roy D.  
REGISTRATION NUMBER: 30,777  
REFERENCE/DOCKET NUMBER: 186991A  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (908) 594-4678  
TELEFAX: (908) 594-4720

INFORMATION FOR SEQ ID NO: 23:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 332 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-560-231-23

Query Match 47.0%; Score 39; DB 2; Length 332;  
Best Local Similarity 56.2%; Pred. No. 72;  
Matches 9; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

OY 3 ATSLAALAVAGDAL 18  
1 : |||: |||: |  
Db 12 ALELVIALSVAGNVL 27

RESULT 11  
US-09-080-704A-23  
Sequence 23, Application US/09080704A  
Patent No. 6166181

GENERAL INFORMATION:  
APPLICANT: Jacobson, Marlene A  
APPLICANT: Johnson, Robert G  
APPLICANT: Luneau, Christopher J  
APPLICANT: Salvatore, Christopher A  
TITLE OF INVENTION: Human Adenosine Receptors  
NUMBER OF SEQUENCES: 28  
CORRESPONDENCE ADDRESS:

ADDRESSEE: Merck & Co., Inc.  
STREET: P.O. Box 2000  
CITY: Rahway  
STATE: NJ  
COUNTRY: United States  
ZIP: 07065

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: Windows NT

GENERAL INFORMATION:  
APPLICANT: Mohlleben, Wolfgang  
APPLICANT: Muth, Gunter  
APPLICANT: Publer, Alfred  
APPLICANT: Rieb, Gunther J.  
TITLE OF INVENTION: Regulated Gene Expression in  
NUMBER OF SEQUENCES: 3  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Finegan, Henderson, Farabow, Garrett &  
STREET: 1300 I Street, N.W., Suite 700  
CITY: Washington  
STATE: D.C.  
COUNTRY: USA  
ZIP: 20005-3315  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/095,481  
FILING DATE: 23-JUL-1993  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/683,062  
FILING DATE: 10-APR-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: DE P 40011863.0  
FILING DATE: 12-APR-1990  
ATTORNEY/AGENT INFORMATION:  
NAME: Madler, Linda A.  
REGISTRATION NUMBER: 33,218  
REFERENCE/DOCKET NUMBER: 02481-1062-01000  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202-408-4000  
TELEFAX: 202-408-4400  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 25 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
STRANDEDNESS: single  
MOLECULE TYPE: peptide  
US-08-095-481-3

Query Match 48.2%; Score 40; DB 1; Length 25;  
Best Local Similarity 57.9%; Pred. No. 2.6;  
Matches 11; Conservative 1; Mismatches 7; Indels 0; Gaps 0;

QY 1 MRATSLAALAVAGDALA 19  
Db 1 MRYRALRLAALVAGAGALA 19

RESULT 6  
5171840-6  
PATENT No. 5171840  
APPLICANT: KISHIMOTO, TADAMITSU  
TITLE OF INVENTION: RECEPTOR PROTEIN FOR HUMAN B CELL  
STIMULATORY FACTOR-2  
NUMBER OF SEQUENCES: 11  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/07/298,694  
FILING DATE: 19-JAN-1989  
SEQ ID NO: 6:  
LENGTH: 323

Query Match 47.0%; Score 39; DB 6; Length 323;  
Best Local Similarity 66.7%; Pred. No. 70;  
Matches 10; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 5 SLAALAVAGDALA 19  
Db 7 ALAALLAAGGALA 21

RESULT 7  
5480796-6  
PATENT No. 5480796  
APPLICANT: KISHIMOTO, TADAMITSU  
TITLE OF INVENTION: ANTIBODIES AGAINST THE RECEPTOR PROTEIN  
FOR HUMAN B CELL STIMULATORY FACTOR-2  
NUMBER OF SEQUENCES: 8  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/07/907,650  
FILING DATE: 02-JUL-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 298,694  
FILING DATE: 19-JAN-1989  
SEQ ID NO: 6:  
LENGTH: 323

Query Match 47.0%; Score 39; DB 6; Length 323;  
Best Local Similarity 66.7%; Pred. No. 70;  
Matches 10; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 5 SLAALAVAGDALA 19  
Db 7 ALAALLAAGGALA 21

RESULT 8  
US-08-349-696-23  
SEQUENCE 23, Application US/08349696  
PATENT No. 559671  
GENERAL INFORMATION:  
APPLICANT: Jacobson, Marlene A  
APPLICANT: Johnson, Robert G  
APPLICANT: Luneau, Christopher J  
APPLICANT: Salvatore, Christopher A  
TITLE OF INVENTION: Human Adenosine Receptors  
NUMBER OF SEQUENCES: 28  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Merck & Co., Inc.  
STREET: P.O. Box 2000  
CITY: Rahway  
STATE: NJ  
COUNTRY: United States  
ZIP: 07065  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: Macintosh itcl  
OPERATING SYSTEM: Macintosh  
SOFTWARE: Microsoft Word 5.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/349,696  
FILING DATE:  
CLASSIFICATION: 530  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: us/08/005945  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Meredith, Roy D.  
REGISTRATION NUMBER: 30,777  
REFERENCE/DOCKET NUMBER: 186991A  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (908)594-4678  
TELEFAX: (908)594-4720  
INFORMATION FOR SEQ ID NO: 23:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 332 amino acids  
TYPE: amino acid

ZIP: 20005  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/07/612,674  
FILING DATE: 19901114  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: SCOTT, WATSON T.  
REGISTRATION NUMBER: 26,581  
REFERENCE/DOCKET NUMBER: 5683/82332  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202-861-3027  
TELEFAX: 202-822-0944  
TELEX: 6714627 CUSH  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 272 amino acids  
TYPE: AMINO ACID  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-07-612-674-2

Query Match 51.8%; Score 43; DB 1; Length 272;  
Best Local Similarity 52.9%; Pred. No. 13;  
Matches 9; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

OY 2 RATSLLAAALAVAGDAL 18  
Db 219 KAELIANSLATAGDGL 235

RESULT 3  
US-07-612-674-5  
Sequence 5, Application US/07612674  
Patent No. 5658792  
GENERAL INFORMATION:  
APPLICANT: NUEL, MARK J.  
APPLICANT: MCCLUNG, J. KEITH  
APPLICANT: STEWART, DAVID A.  
TITLE OF INVENTION: AN ANTIPROLIFERATIVE PROTEIN  
NUMBER OF SEQUENCES: 14  
CORRESPONDENCE ADDRESSES:  
ADDRESSEE: CUSHMAN, DARBY & CUSHMAN  
STREET: 1100 NEW YORK AVENUE, N.W.  
CITY: WASHINGTON  
STATE: D.C.  
COUNTRY: USA  
ZIP: 20005  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/07/612,674  
FILING DATE: 19901114  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: SCOTT, WATSON T.  
REGISTRATION NUMBER: 26,581  
REFERENCE/DOCKET NUMBER: 5683/82332  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202-861-3027  
TELEFAX: 202-822-0944  
TELEX: 6714627 CUSH  
INFORMATION FOR SEQ ID NO: 5:  
SEQUENCE CHARACTERISTICS:

LENGTH: 276 amino acids  
TYPE: AMINO ACID  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-07-612-674-5

Query Match 50.6%; Score 42; DB 1; Length 276;  
Best Local Similarity 62.5%; Pred. No. 19;  
Matches 10; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

OY 3 ATSLAAALAVAGDAL 18  
Db 220 AAGLAASLAERAGDGL 235

RESULT 4  
US-08-293-563-5  
Sequence 5, Application US/08293563  
Patent No. 5516894  
GENERAL INFORMATION:  
APPLICANT: Reperit, Steven M.  
TITLE OF INVENTION: A2b-ADENOSINE RECEPTORS AND  
NUMBER OF SEQUENCES: 7  
RELATED MOLECULES AND METHODS  
CORRESPONDENCE ADDRESSES:  
ADDRESSEE: Fish & Richardson  
STREET: 225 Franklin Street  
CITY: Boston  
STATE: Massachusetts  
COUNTRY: U.S.A.  
ZIP: 02110-2804  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 3.5" Diskette, 1.44 Mb  
COMPUTER: IBM PS/2 Model 50Z or 55SX  
OPERATING SYSTEM: MS-DOS (Version 5.0)  
SOFTWARE: Wordperfect (Version 5.1)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/293,563  
FILING DATE:  
CLASSIFICATION: 530  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 07/850,188  
FILING DATE: March 11, 1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Clark, Paul T.  
REGISTRATION NUMBER: 30,162  
REFERENCE/DOCKET NUMBER: 00786/128001  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617) 542-5070  
TELEFAX: (617) 542-8906  
TELEX: 200154  
INFORMATION FOR SEQ ID NO: 5:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 332  
TYPE: amino acid  
STRANDEDNESS: N/A  
TOPOLOGY: linear  
US-08-293-563-5

Query Match 50.6%; Score 42; DB 1; Length 332;  
Best Local Similarity 62.5%; Pred. No. 24;  
Matches 10; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

OY 3 ATSLAAALAVAGDAL 18  
Db 12 ALELVIAALAVAGNVL 27

RESULT 5  
US-08-095-481-3  
Sequence 3, Application US/08095481  
Patent No. 5336607

GenCore version 5.1.4\_p5-4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 9, 2003, 14:57:25 ; Search time 1.17123 Seconds  
(without alignments)  
477.305 Million cell updates/sec

Title: US-10-028-245-3

Perfect score: 83

Sequence: 1 MRATSLAALAVAGDALA 19

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 262574 segs, 29422922 residues

262574

Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued\_Patents\_AA:\*

1: /cgn2\_6/ptodata/1/1aa/5A.COMB.pep:\*

2: /cgn2\_6/ptodata/1/1aa/5B.COMB.pep:\*

3: /cgn2\_6/ptodata/1/1aa/6A.COMB.pep:\*

4: /cgn2\_6/ptodata/1/1aa/6B.COMB.pep:\*

5: /cgn2\_6/ptodata/1/1aa/6C.COMB.pep:\*

6: /cgn2\_6/ptodata/1/1aa/6D.COMB.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	44	53.0	578	4 US-09-503-172A-2	Sequence 2, Appl
2	43	51.8	272	1 US-07-612-674-2	Sequence 2, Appl
3	42	50.6	276	1 US-07-612-674-5	Sequence 5, Appl
4	40	48.2	25	1 US-08-095-481-3	Sequence 3, Appl
5	39	47.0	323	6 5171840-6	Patent No. 5171840
6	39	47.0	323	6 5460796-6	Patent No. 5460796
7	39	47.0	332	1 US-08-349-696-23	Sequence 23, Appl
8	39	47.0	332	1 US-08-233-009-23	Sequence 23, Appl
9	39	47.0	332	2 US-08-560-231-23	Sequence 23, Appl
10	39	47.0	332	2 US-08-560-231-23	Sequence 23, Appl
11	39	47.0	332	2 US-08-560-231-23	Sequence 23, Appl
12	39	47.0	332	2 US-08-560-231-23	Sequence 23, Appl
13	39	47.0	332	2 US-08-560-231-23	Sequence 23, Appl
14	39	47.0	332	2 US-08-560-231-23	Sequence 23, Appl
15	39	47.0	332	2 US-08-560-231-23	Sequence 23, Appl
16	39	47.0	332	2 US-08-560-231-23	Sequence 23, Appl
17	39	47.0	332	2 US-08-560-231-23	Sequence 23, Appl
18	39	47.0	332	2 US-08-560-231-23	Sequence 23, Appl
19	39	47.0	332	2 US-08-560-231-23	Sequence 23, Appl
20	39	47.0	332	2 US-08-560-231-23	Sequence 23, Appl
21	39	47.0	332	2 US-08-560-231-23	Sequence 23, Appl
22	39	47.0	332	2 US-08-560-231-23	Sequence 23, Appl
23	39	47.0	332	2 US-08-560-231-23	Sequence 23, Appl
24	39	47.0	332	2 US-08-560-231-23	Sequence 23, Appl
25	39	47.0	332	2 US-08-560-231-23	Sequence 23, Appl
26	39	47.0	332	2 US-08-560-231-23	Sequence 23, Appl
27	39	47.0	332	2 US-08-560-231-23	Sequence 23, Appl

28	39	47.0	1832	4 US-09-567-969-4	Sequence 4, Appl
29	39	47.0	1832	4 US-09-568-480-4	Sequence 4, Appl
30	39	47.0	1832	4 US-09-568-486-4	Sequence 4, Appl
31	39	47.0	1832	4 US-09-568-472-4	Sequence 4, Appl
32	39	47.0	1832	4 US-09-567-899-4	Sequence 4, Appl
33	39	47.0	1832	4 US-08-399-696-70	Sequence 70, Appl
34	39	47.0	1832	4 US-08-158-189-14	Sequence 14, Appl
35	39	47.0	1832	4 US-08-158-189-15	Sequence 15, Appl
36	39	47.0	1832	4 US-08-399-696-122	Sequence 122, App
37	39	47.0	1832	4 US-08-399-696-102	Sequence 102, App
38	39	47.0	1832	4 US-08-545-860D-55	Sequence 55, Appl
39	39	47.0	1832	4 US-08-545-860D-55	Sequence 55, Appl
40	39	47.0	1832	4 US-08-545-860D-55	Sequence 55, Appl
41	39	47.0	1832	4 US-08-545-860D-55	Sequence 55, Appl
42	39	47.0	1832	4 US-08-545-860D-55	Sequence 55, Appl
43	39	47.0	1832	4 US-08-545-860D-55	Sequence 55, Appl
44	39	47.0	1832	4 US-08-545-860D-55	Sequence 55, Appl
45	39	47.0	1832	4 US-08-545-860D-55	Sequence 55, Appl

## ALIGNMENTS

RESULT 1

US-09-503-172A-2

Sequence 2, Application US/09503172A

Patent No. 6284510

GENERAL INFORMATION:

APPLICANT: ITO, Tetsuya

APPLICANT: FUJITA, Koki

APPLICANT: HARA, Kozo

APPLICANT: TONOKURA, Takashi

APPLICANT: SAKANO, Yoshiyuki

TITLE OF INVENTION: BETA-PRCTUFURANOSIDASE GENE

FILE REFERENCE: 10749-0001-0

CURRENT FILING DATE: 2000-02-14

PRIOR APPLICATION NUMBER: JP 160416/1999

PRIOR FILING DATE: 1999-06-08

NUMBER OF SEQ ID NOS: 9

SOFTWARE: Patentin Ver. 2.1

SEQ ID NO 2

LENGTH: 578

TYPE: PRT

ORGANISM: Arthrobacter sp.

US-09-503-172A-2

Query Match

Best Local Similarity 53.0%; Score 44; DB 4; Length 578;

Matches 11; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 3 ATSLAALAVAGDALA 19

DB 18 ATSLAALAVAGDALA 34

RESULT 2

US-07-612-674-2

Sequence 2, Application US/07612674

Patent No. 5658792

GENERAL INFORMATION:

APPLICANT: NIELL, MARK J.

APPLICANT: MCGILL, J. KEITH

APPLICANT: STEWART, DAVID A.

APPLICANT: DANNER, DAVID B.

TITLE OF INVENTION: AN ANTIPROLIFERATIVE PROTEIN

NUMBER OF SEQUENCES: 14

CORRESPONDENCE ADDRESS:

ADDRESSEE: CUSHMAN, DARBY & CUSHMAN

STREET: 1100 NEW YORK AVENUE, N.W.

CITY: WASHINGTON

STATE: D.C.

COUNTRY: USA

**THIS PAGE BLANK (USPTO)**

DB 1 MRATLTLALLALALCIACGA 21

## RESULT 15

AAU10688  
ID AAU10688 standard; Protein; 100 AA.

AC AAU10688;

DT 14-FEB-2002 (first entry)

DE Polymorphic variant of human BGLAP protein.

Human; single nucleotide polymorphism; SNP; BGLAP; chromosome 1q25-q31;  
bone gamma carboxyglutamate protein; haplotyping; genotyping;  
osteoporosis; osteopathic; variant.

OS Homo sapiens.

Key Location/Qualifiers

FT Misc-difference 82 /note= "Substitution of Glu to Lys"

FT Misc-difference 94 /note= "Substitution of Arg to Gln"

PN WO20017131-A2.

PD 18-OCT-2001.

PF 11-APR-2001; 2001WO-US12010.

PR 11-APR-2000; 2000US-195840P.

PA (GENA-) GENAISSANCE PHARM INC.

PI Bentivegna SC, Chew A, Choi JY, Koshy B, Rounds E, Stephens JC;

DR WPI; 2002-041288/05.

PT New haplotypes of the human bone gamma carboxyglutamate protein gene,  
useful to diagnose and treat diseases associated with the gene such as  
osteoporosis

PS Claim 27; Page -: 53pp; English.

CC The present invention relates to novel single nucleotide polymorphisms  
(SNPs) in the human bone gamma carboxyglutamate protein (BGLAP) gene  
located on chromosome 1q25-q31, and methods for haplotyping and/or  
genotyping the BGLAP gene in an individual. The methods of the  
invention make use of allele-specific oligonucleotides (ASOs) as probes  
and primers and/or primer-extension oligonucleotides for detecting the  
BGLAP gene polymorphisms. The polynucleotides and screened compounds are  
useful for (developing) treatment of diseases associated with BGLAP  
activity, such as osteoporosis. The present sequence represents a  
polymorphic variant of the BGLAP protein (AAU10687).  
CC Note: The present sequence is not given in the specification but is  
created by the indexer from the information given in the patent.

XX SQ Sequence 100 AA;

Query Match 51.8%; Score 43; DB 23; Length 100;

Best Local Similarity 57.1%; Pred. No. 21;  
Matches 12; Conservative 2; Mismatches 4; Indels 3; Gaps 1;

OY 1 MRATSLA---AALAVAGDA 17

DB 1 MRATLTLALLALALCIACGA 21

Search completed: May 9, 2003, 15:18:32  
Job time : 4.81963 secs

PT increased heat resistance and transfer ratio by means of genetic  
 PT engineering techniques -  
 XX  
 XX  
 PS Disclosure; Page 24; 34pp; English.  
 XX  
 CC The present invention relates a beta-fructofuranosidase gene.  
 CC The invention is useful for the development of variant enzymes  
 CC that have increased heat resistance and transfer ratio by means  
 CC of genetic engineering techniques. Beta-fructofuranosidase is  
 CC useful in the synthesis of transfructosylated oligosaccharides  
 CC such as lactosucrose, for use in fields of foods and drugs.  
 CC  
 XX  
 SQ Sequence 578 AA;  
 OY  
 Db 3 ATSLAALAVAGDAA 19  
 ||| ||||: ||| |  
 18 ATSLAALAVLAGAAPA 34  
 Query Match 53.0%; Score 44; DB 22; Length 578;  
 Best Local Similarity 64.7%; Pred. No. 98;  
 Matches 11; Conservative 2; Mismatches 4; Indels 0; Gaps 0;  
 RESULT 13  
 AAW76094  
 ID AAW76094 standard; Protein; 98 AA.  
 AC AAW76094;  
 XX  
 DT 21-DEC-1998 (first entry)  
 XX  
 DE Human osteocalcin protein.  
 XX  
 KW Promoter; tissue-specific gene expression; skeletal tissue;  
 KW stem cell; bone; cartilage; osteocalcin; hOC gene; human;  
 KW osteoporosis; osteopenia; osteosarcoma; cancer; metastasis;  
 KW gene therapy.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WC9839427-A2.  
 PD 11-SEP-1998.  
 XX  
 PF 06-MAR-1998; 98MO-US04421.  
 XX  
 PR 06-MAR-1997; 97US-0039839.  
 XX  
 PA (UYMA-) UNIV MASSACHUSETTS.  
 XX  
 PI Frenkel B, Hou Z, Lian JB, Nilsson S, Peters S;  
 PI Quisenberry P, Stein GS, Stein JI;  
 XX  
 DR WPI; 1998-495839/42.  
 DR N-PSDB; AAV46429.  
 XX  
 PT Expression of exogenous genes in differentiated cells - by  
 PT transducing pluripotent stem cells capable of maturing into  
 PT differentiated cells with nucleic acid comprising exogenous gene,  
 PT useful for, e.g. treatment of osteoporosis  
 XX  
 PS Disclosure; Page 33-34; 63pp; English.  
 XX  
 CC This is the amino acid sequence of the human osteocalcin, deduced  
 CC from the coding exons of the human hOC gene (see AAV46429). The  
 CC invention pertains to a method for expressing endogenous genes in  
 CC differentiated cells of a specific type. The method involves  
 CC contacting pluripotent stem cells capable of maturing into  
 CC differentiated cells with a nucleic acid comprising an exogenous  
 CC gene linked to a regulatory element capable of controlling expression  
 CC of the exogenous gene in the differentiated cells. A population of  
 CC transduced stem cells capable of maturing into differentiated cells  
 CC expressing the exogenous gene is produced. Preferably, the

CC differentiated cells are in a tissue of interest, such as bone or  
 CC cartilage, and the exogenous gene is operably linked to at least one  
 CC osteocalcin regulatory element derived from the hOC promoter. The  
 CC exogenous gene can encode a therapeutic protein useful for treating  
 CC a disease, especially osteoporosis, osteopenia, osteosarcoma,  
 CC primary malignancy or metastases (all claimed).  
 CC  
 XX  
 SQ Sequence 98 AA;  
 OY  
 Db 1 MRATSLA----AALAVAGDA 17  
 ||| |||| ||| ||| |  
 1 MRATSLAALLAALNCINGQA 21  
 Query Match 51.8%; Score 43; DB 19; Length 98;  
 Best Local Similarity 57.1%; Pred. No. 21;  
 Matches 12; Conservative 2; Mismatches 3; Indels 4; Gaps 1;  
 RESULT 14  
 AAU10687  
 ID AAU10687 standard; Protein; 100 AA.  
 AC AAU10687;  
 XX  
 DT 14-FEB-2002 (first entry)  
 XX  
 DE Reference sequence for human BGLAP protein.  
 XX  
 KW Human; single nucleotide polymorphism; SNP; BGLAP; chromosome 1q25-q31;  
 KW bone gamma carboxyglutamate protein; haplotyping; genotyping;  
 KW osteoporosis; osteopathic.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WC200177131-A2.  
 PD 18-OCT-2001.  
 XX  
 PF 11-APR-2001; 2001WO-US12010.  
 XX  
 PR 11-APR-2000; 2000US-195840P.  
 XX  
 PA (GENA-) GENAISSANCE PHARM INC.  
 XX  
 PI Bentivegna SC, Chew A, Choi JY, Koshy B, Rounds E, Stephens JC;  
 PI WPI; 2002-041288/05.  
 DR N-PSDB; AAS16363, AAS16364.  
 XX  
 PT New haplotypes of the human bone gamma carboxyglutamate protein gene,  
 PT useful to diagnose and treat diseases associated with the gene such as  
 PT osteoporosis -  
 XX  
 PS Claim 27; Fig 3; 53pp; English.  
 XX  
 CC The present invention relates to novel single nucleotide polymorphisms  
 CC (SNPs) in the human bone gamma carboxyglutamate protein (BGLAP) gene  
 CC located on chromosome 1q25-q31, and methods for haplotyping and/or  
 CC genotyping the BGLAP gene in an individual. The methods of the  
 CC invention make use of allele-specific oligonucleotides (ASOs) as probes  
 CC and primers and/or primer-extension oligonucleotides (ASOs) as probes  
 CC BGLAP gene polymorphisms. The polymorphisms and screened compounds are  
 CC useful for (developing) treatment of diseases associated with BGLAP  
 CC activity, such as osteoporosis. The present sequence represents a  
 CC reference sequence for the BGLAP protein.  
 CC  
 XX  
 SQ Sequence 100 AA;  
 OY  
 Db 1 MRATSLA----AALAVAGDA 17  
 ||| |||| ||| ||| |  
 1 MRATSLAALLAALNCINGQA 21  
 Query Match 51.8%; Score 43; DB 23; Length 100;  
 Best Local Similarity 57.1%; Pred. No. 21;  
 Matches 12; Conservative 2; Mismatches 3; Indels 4; Gaps 1;

XX Claim 6; SEQ ID NO 1160; 2504pp; French.

PT The present invention is related to a Lactococcus lactis nucleotide  
XX sequence (ABA050521) and related proteins (ABB53300-ABB55621). The  
CC nucleic acid sequence is useful in the detection and/or amplification of  
CC nucleic acid sequence, particularly to identify Lactococcus lactis or  
CC related species. The proteins of the invention are useful for the  
CC biosynthesis or biodegradation of a composition of interest. The  
CC invention helps research in lactic bacteria, particularly useful in the  
CC production of yogurt and cheese.

CC Note: The sequence data for this patent is based on equivalent patent  
CC WO2001/7734 (published 18-OCT-2001) which is available in electronic  
CC format directly from WIPO at ftp.wipo.int/pub/published\_pct\_sequences.

XX  
XX  
SQ Sequence 166 AA:

Query Match 54.2%; Score 45; DB 23; Length 166;  
Best Local Similarity 56.2%; Pred. No. 18;  
Matches 9; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

0Y 1 MRATSLAAALAVAGD 16  
Db 22 MRISLLAVALQISGE 37

RESULT 11  
AAG68270  
ID AAG68270 standard; Protein: 271 AA.

XX AAG68270;  
XX  
XX 14-FEB-2002 (first entry)

DE Human. POLY17 protein sequence SEQ ID NO:34.

XX Human; POLYX; gamma aminobutyric acid receptor; GABA receptor;  
KW epidermal growth factor; EGF; complement receptor; HSPC; syntaxin;  
KW haematopoietic stem and progenitor cell; sulphotransferase; prolactin;  
KW antidepressant; cerebroprotective; antiparkinsonian; nootropic; relaxant;  
KW anticonvulsant; neuroleptic; neuroprotective; antialcoholic; cardiant;  
KW tranquilliser; antiarrhythmic; psychiatric; medical; depression; stroke;  
KW Parkinson's disease; Huntington's disease; Tourette's syndrome; anxiety;  
KW amyotrophic lateral sclerosis; head trauma; Alzheimer's disease;  
KW alcoholism; vigilance; muscle tension; epileptogenic; memory function;  
KW cardiomyopathy; arrhythmogenic right ventricular dysplasia.

XX  
XX Homo sapiens.  
XX  
XX WO200179294-A2.  
XX  
XX 25-OCT-2001.

PD 19-APR-2001; 2001WO-US1854.  
XX  
XX 19-APR-2000; 2000US-198293P.  
XX 20-APR-2000; 2000US-198645P.  
XX 25-APR-2000; 2000US-199476P.  
XX 26-APR-2000; 2000US-199880P.  
XX 26-APR-2000; 2000US-200024P.  
XX 26-APR-2000; 2000US-200025P.  
XX 09-JUN-2000; 2000US-210809P.  
XX 17-JUL-2000; 2000US-218591P.  
XX 11-AUG-2000; 2000US-224610P.  
XX 09-FEB-2001; 2001US-267673P.

XX  
XX (CURA-) CURAGEN CORP.  
XX  
XX Taupier RJ, Vernet CM, Fernandes E, Shinkens RA, Majumder K;  
PI Padigaru M, Colman SD, Zernhusen BD, Szytek KA, Burgess CE, Liu X;  
XX WPI: 2002-017601/02.  
XX N-PSDB; ABA03886.

XX  
XX  
PT New isolated polypeptides for treating a broad range of pathological  
PT states, e.g., depression, stroke, Parkinson's disease, Huntington's  
PT disease, Tourette's syndrome, amyotrophic lateral sclerosis, head  
PT trauma, and Alzheimer's

XX  
XX  
PS Claim 1; Page 64; 155pp; English.

CC The present invention describes polypeptides (I), designated POLYX  
CC polypeptides (e.g. POLY1, POLY2, etc.), and the polynucleotide  
CC sequences (II) encoding them. POLY1-4 are members of the gamma  
CC aminobutyric acid (GABA) receptor family; POLY5-8 are members of the  
CC epidermal growth factor (EGF) family; POLY9-11 are members of the  
CC complement receptor family; POLY12 is a member of the haematopoietic  
CC stem and progenitor cell (HSPC) family; POLY13 is a member of the  
CC sulphotransferase family; POLY14-16 are members of the syntaxin family;  
CC and POLY17 is a member of the prolactin family. (I) and (II) can have  
CC antidepressant, cerebroprotective, antiparkinsonian, nootropic, relaxant,  
CC anticonvulsant, neuroleptic, neuroprotective, antialcoholic, cardiant,  
CC tranquilliser and antiarrhythmic activities. (I) and (II) can be used  
CC for treating or preventing a POLYX-associated disorder in humans as a  
CC therapeutic in the manufacture of a medicament for treating a syndrome  
CC associated with a human disease selected from a POLYX-associated  
CC disorder, for treating a pathological state in a mammal, especially  
CC patients suffering from, e.g., psychiatric and medical conditions,  
CC depression, stroke, Parkinson's disease, Huntington's disease,  
CC Tourette's syndrome, amyotrophic lateral sclerosis, head trauma,  
CC Alzheimer's disease, alcoholism, vigilance, anxiety, muscle tension,  
CC epileptogenic activity and memory functions, cardiomyopathy and  
CC arrhythmogenic right ventricular dysplasia. The present sequence  
XX represents POLY17.

XX  
XX  
SQ Sequence 271 AA:

Query Match 53.0%; Score 44; DB 23; Length 271;  
Best Local Similarity 47.1%; Pred. No. 43;  
Matches 8; Conservative 5; Mismatches 4; Indels 0; Gaps 0;

0Y 2 RATSLAAALAVAGDAL 18  
Db 214 KATFPAASVATMGDGL 230

RESULT 12  
AAB66692  
ID AAB66692 standard; Protein: 578 AA.

XX AAB66692;  
XX  
XX 06-APR-2001 (first entry)

DE Beta-fructofuranosidase protein #1.

XX Beta-fructofuranosidase; transfructosylated; food; drug.  
XX  
XX Arthrobacter sp.  
XX  
XX CA2298400-A1.  
XX  
XX 08-DEC-2000.  
XX  
XX 17-FEB-2000; 2000CA-2298400.  
XX  
XX 08-JUN-1999; 99JP-0160416.  
XX  
XX (NORO) SOC TECHNO-INNOVATION AGRIC FORESTY & FL.

XX Hara K, Tonozuka T, Ito T, Sakano Y, Fujita K;  
PI WPI: 2001-123506/14.  
XX  
XX Novel beta fructofuranosidase gene useful for producing beta  
XX fructofuranosidase and for developing variant enzymes that have

PR 14-JUL-1999; 99US-0143624.  
 PR 15-JUL-1999; 99US-0144005.  
 PR 16-JUL-1999; 99US-0144085.  
 PR 16-JUL-1999; 99US-0144086.  
 PR 19-JUL-1999; 99US-0144325.  
 PR 19-JUL-1999; 99US-0144331.  
 PR 19-JUL-1999; 99US-0144332.  
 PR 19-JUL-1999; 99US-0144333.  
 PR 19-JUL-1999; 99US-0144334.  
 PR 19-JUL-1999; 99US-0144335.  
 PR 20-JUL-1999; 99US-0144352.  
 PR 20-JUL-1999; 99US-0144632.  
 PR 20-JUL-1999; 99US-0144684.  
 PR 21-JUL-1999; 99US-0144814.  
 PR 21-JUL-1999; 99US-0145086.  
 PR 21-JUL-1999; 99US-0145088.  
 PR 22-JUL-1999; 99US-0145085.  
 PR 22-JUL-1999; 99US-0145087.  
 PR 22-JUL-1999; 99US-0145089.  
 PR 22-JUL-1999; 99US-0145192.  
 PR 23-JUL-1999; 99US-0145145.  
 PR 23-JUL-1999; 99US-0145218.  
 PR 23-JUL-1999; 99US-0145224.  
 PR 26-JUL-1999; 99US-0145276.  
 PR 27-JUL-1999; 99US-0145913.  
 PR 27-JUL-1999; 99US-0145918.  
 PR 28-JUL-1999; 99US-0145919.  
 PR 02-AUG-1999; 99US-0145951.  
 PR 02-AUG-1999; 99US-0146386.  
 PR 02-AUG-1999; 99US-0146388.  
 PR 02-AUG-1999; 99US-0146389.  
 PR 03-AUG-1999; 99US-0147038.  
 PR 04-AUG-1999; 99US-0147204.  
 PR 04-AUG-1999; 99US-0147302.  
 PR 05-AUG-1999; 99US-0147192.  
 PR 05-AUG-1999; 99US-0147260.  
 PR 06-AUG-1999; 99US-0147303.  
 PR 06-AUG-1999; 99US-0147416.  
 PR 09-AUG-1999; 99US-0147493.  
 PR 10-AUG-1999; 99US-0148171.  
 PR 11-AUG-1999; 99US-0148319.  
 PR 12-AUG-1999; 99US-0148341.  
 PR 13-AUG-1999; 99US-0148565.  
 PR 13-AUG-1999; 99US-0148565.  
 PR 16-AUG-1999; 99US-0149368.  
 PR 17-AUG-1999; 99US-0149175.  
 PR 18-AUG-1999; 99US-0149426.  
 PR 20-AUG-1999; 99US-0149722.  
 PR 20-AUG-1999; 99US-0149723.  
 PR 20-AUG-1999; 99US-0149929.  
 PR 23-AUG-1999; 99US-0149902.  
 PR 23-AUG-1999; 99US-0149930.  
 PR 25-AUG-1999; 99US-0150566.  
 PR 26-AUG-1999; 99US-0150884.  
 PR 27-AUG-1999; 99US-0151065.  
 PR 27-AUG-1999; 99US-0151086.  
 PR 30-AUG-1999; 99US-0151303.  
 PR 31-AUG-1999; 99US-0151438.  
 PR 01-SEP-1999; 99US-0151930.  
 PR 07-SEP-1999; 99US-0152363.  
 PR 10-SEP-1999; 99US-0153070.  
 PR 13-SEP-1999; 99US-0153758.  
 PR 15-SEP-1999; 99US-0154018.  
 PR 16-SEP-1999; 99US-0154039.  
 PR 20-SEP-1999; 99US-0154779.  
 PR 22-SEP-1999; 99US-0155139.  
 PR 23-SEP-1999; 99US-0155486.  
 PR 24-SEP-1999; 99US-0155659.  
 PR 28-SEP-1999; 99US-0156458.  
 PR 29-SEP-1999; 99US-0156596.  
 PR 04-OCT-1999; 99US-0157117.

PR 05-OCT-1999; 99US-0157753.  
 PR 06-OCT-1999; 99US-0157865.  
 PR 07-OCT-1999; 99US-0158029.  
 PR 08-OCT-1999; 99US-0158232.  
 PR 12-OCT-1999; 99US-0158359.  
 PR 13-OCT-1999; 99US-0159293.  
 PR 13-OCT-1999; 99US-0159294.  
 PR 13-OCT-1999; 99US-0159295.  
 PR 14-OCT-1999; 99US-0159329.  
 PR 14-OCT-1999; 99US-0159330.  
 PR 14-OCT-1999; 99US-0159331.  
 PR 14-OCT-1999; 99US-0159637.  
 PR 14-OCT-1999; 99US-0159638.  
 PR 18-OCT-1999; 99US-0159584.  
 PR 21-OCT-1999; 99US-0160741.  
 PR 21-OCT-1999; 99US-0160767.  
 PR 21-OCT-1999; 99US-0160768.  
 PR 21-OCT-1999; 99US-0160770.  
 PR 21-OCT-1999; 99US-0160814.  
 PR 21-OCT-1999; 99US-0160815.  
 PR 22-OCT-1999; 99US-0160980.  
 PR 22-OCT-1999; 99US-0160981.  
 PR 22-OCT-1999; 99US-0160989.  
 PR 25-OCT-1999; 99US-0161404.  
 PR 25-OCT-1999; 99US-0161405.  
 PR 25-OCT-1999; 99US-0161406.  
 PR 26-OCT-1999; 99US-0161359.  
 PR 26-OCT-1999; 99US-0161360.  
 PR 26-OCT-1999; 99US-0161361.  
 PR 28-OCT-1999; 99US-0161920.  
 PR 28-OCT-1999; 99US-0161992.  
 PR 28-OCT-1999; 99US-0161993.  
 PR 29-OCT-1999; 99US-0162142.

Query Match 55.4%; Score 46; DB 21; Length 515;  
 Best Local Similarity 68.8%; Pred. No. 42;  
 Matches 11; Conservative 1; Mismatches 4; Indels 0; Gaps 0;  
 Oy 3 ATSLAALAVAGDAL 18  
 Db 353 ATSLADGYAVAGQAI 368

RESULT 10  
 ABB54458  
 ID ABB54458 standard; Protein: 166 AA.  
 AC ABB54458;  
 DT 16-MAY-2002 (first entry)  
 XX  
 DE Lactococcus lactis protein yllf.  
 XX  
 KW Biosynthesis; biodegradation; lactic bacterium; yogurt; cheese.  
 OS Lactococcus lactis IL1403.  
 PN FR2807446-A1.  
 PD 12-OCT-2001.  
 PF 11-APR-2000; 2000FR-0004630.  
 PR 11-APR-2000; 2000FR-0004630.  
 XX  
 PA (INRG ) INRA INST NAT RECH AGRONOMICUE.  
 XX Bolotine A, Sorokine A, Renault P, Ehrlich SD;  
 PI WPI; 2002-043418/06.  
 DR  
 XX New nucleotide sequence useful in the identification or Lactococcus  
 PT lactic and related species -

PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156658.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161992.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.  
  
Query Match 55.48; Score 46; DB 21; Length 514;  
Best Local Similarity 68.8%; Pred. No. 42;  
Matches 11; Conservative 1; Mismatches 4; Indels 0; Gaps 0;  
  
OY 3 ATSLAALAVAGDAL 18  
DB 352 ATSLADGYAVAGQAI 367  
  
RESULT 9  
AAG42079  
ID AAG42079 standard; Protein; 515 AA.  
XX  
AC AAG42079;  
XX  
DT 18-OCT-2000 (first entry)  
XX  
DE Arabidopsis thaliana protein fragment SEQ ID NO: 52434.  
XX  
KW Protein identification; signal transduction pathway; metabolic pathway;  
hybridisation assay; genetic mapping; gene expression control; promoter;  
termination sequence.  
XX  
OS Arabidopsis thaliana.  
XX  
PN EP1033405-A2.  
XX  
PD 06-SEP-2000.  
XX  
PF 25-FEB-2000; 2000EP-0301439.  
XX  
PR 25-FEB-1999; 99US-0121825.

PR 05-MAR-1999; 99US-0123180.  
PR 09-MAR-1999; 99US-0123548.  
PR 23-MAR-1999; 99US-0125788.  
PR 25-MAR-1999; 99US-0126264.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 28-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 30-APR-1999; 99US-0132407.  
PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 06-MAY-1999; 99US-0132487.  
PR 07-MAY-1999; 99US-0132853.  
PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.

PN EPI033405-A2.  
 XX 06-SEP-2000.  
 XX 25-FEB-2000; 2000EP-0301439.  
 PF  
 XX  
 PR 25-FEB-1999; 99US-0121825.  
 PR 05-MAR-1999; 99US-0123180.  
 PR 09-MAR-1999; 99US-0123548.  
 PR 23-MAR-1999; 99US-0125788.  
 PR 25-MAR-1999; 99US-0126264.  
 PR 29-MAR-1999; 99US-0126785.  
 PR 01-APR-1999; 99US-0127462.  
 PR 08-APR-1999; 99US-0128234.  
 PR 16-APR-1999; 99US-0129845.  
 PR 19-APR-1999; 99US-0130077.  
 PR 21-APR-1999; 99US-0130449.  
 PR 23-APR-1999; 99US-0130510.  
 PR 28-APR-1999; 99US-0131449.  
 PR 30-APR-1999; 99US-0132048.  
 PR 30-APR-1999; 99US-0132407.  
 PR 04-MAY-1999; 99US-0132484.  
 PR 05-MAY-1999; 99US-0132485.  
 PR 06-MAY-1999; 99US-0132486.  
 PR 06-MAY-1999; 99US-0132487.  
 PR 07-MAY-1999; 99US-0132863.  
 PR 11-MAY-1999; 99US-0134256.  
 PR 14-MAY-1999; 99US-0134218.  
 PR 14-MAY-1999; 99US-0134219.  
 PR 14-MAY-1999; 99US-0134421.  
 PR 18-MAY-1999; 99US-0134370.  
 PR 19-MAY-1999; 99US-0134941.  
 PR 20-MAY-1999; 99US-0135124.  
 PR 21-MAY-1999; 99US-0135353.  
 PR 24-MAY-1999; 99US-0135629.  
 PR 25-MAY-1999; 99US-0136021.  
 PR 27-MAY-1999; 99US-0136392.  
 PR 28-MAY-1999; 99US-0136782.  
 PR 01-JUN-1999; 99US-0137222.  
 PR 03-JUN-1999; 99US-0137528.  
 PR 04-JUN-1999; 99US-0137502.  
 PR 07-JUN-1999; 99US-0137724.  
 PR 08-JUN-1999; 99US-0138094.  
 PR 10-JUN-1999; 99US-0138540.  
 PR 10-JUN-1999; 99US-0138847.  
 PR 14-JUN-1999; 99US-0139119.  
 PR 16-JUN-1999; 99US-0139452.  
 PR 16-JUN-1999; 99US-0139453.  
 PR 17-JUN-1999; 99US-0139492.  
 PR 18-JUN-1999; 99US-0139454.  
 PR 18-JUN-1999; 99US-0139455.  
 PR 18-JUN-1999; 99US-0139456.  
 PR 18-JUN-1999; 99US-0139457.  
 PR 18-JUN-1999; 99US-0139458.  
 PR 18-JUN-1999; 99US-0139459.  
 PR 18-JUN-1999; 99US-0139460.  
 PR 18-JUN-1999; 99US-0139461.  
 PR 18-JUN-1999; 99US-0139462.  
 PR 18-JUN-1999; 99US-0139463.  
 PR 18-JUN-1999; 99US-0139750.  
 PR 18-JUN-1999; 99US-0139763.  
 PR 21-JUN-1999; 99US-0139817.  
 PR 22-JUN-1999; 99US-0139899.  
 PR 23-JUN-1999; 99US-0140353.  
 PR 23-JUN-1999; 99US-0140354.  
 PR 24-JUN-1999; 99US-0140695.  
 PR 28-JUN-1999; 99US-0140823.  
 PR 29-JUN-1999; 99US-0140991.  
 PR 30-JUN-1999; 99US-0141287.  
 PR 01-JUL-1999; 99US-0141842.

PR 01-JUL-1999; 99US-0142154.  
 PR 02-JUL-1999; 99US-0142055.  
 PR 06-JUL-1999; 99US-0142390.  
 PR 08-JUL-1999; 99US-0142803.  
 PR 09-JUL-1999; 99US-0142920.  
 PR 12-JUL-1999; 99US-0142977.  
 PR 13-JUL-1999; 99US-0143542.  
 PR 14-JUL-1999; 99US-0143624.  
 PR 15-JUL-1999; 99US-0144005.  
 PR 16-JUL-1999; 99US-0144085.  
 PR 16-JUL-1999; 99US-0144086.  
 PR 19-JUL-1999; 99US-0144325.  
 PR 19-JUL-1999; 99US-0144331.  
 PR 19-JUL-1999; 99US-0144332.  
 PR 19-JUL-1999; 99US-0144333.  
 PR 19-JUL-1999; 99US-0144334.  
 PR 19-JUL-1999; 99US-0144335.  
 PR 20-JUL-1999; 99US-0144352.  
 PR 20-JUL-1999; 99US-0144632.  
 PR 20-JUL-1999; 99US-0144684.  
 PR 21-JUL-1999; 99US-0144814.  
 PR 21-JUL-1999; 99US-0145086.  
 PR 21-JUL-1999; 99US-0145088.  
 PR 22-JUL-1999; 99US-0145085.  
 PR 22-JUL-1999; 99US-0145087.  
 PR 22-JUL-1999; 99US-0145089.  
 PR 22-JUL-1999; 99US-0145192.  
 PR 22-JUL-1999; 99US-0145145.  
 PR 23-JUL-1999; 99US-0145218.  
 PR 23-JUL-1999; 99US-0145224.  
 PR 26-JUL-1999; 99US-0145276.  
 PR 27-JUL-1999; 99US-0145913.  
 PR 27-JUL-1999; 99US-0145918.  
 PR 28-JUL-1999; 99US-0145951.  
 PR 02-AUG-1999; 99US-0146386.  
 PR 02-AUG-1999; 99US-0146388.  
 PR 02-AUG-1999; 99US-0146389.  
 PR 03-AUG-1999; 99US-0147038.  
 PR 04-AUG-1999; 99US-0147204.  
 PR 04-AUG-1999; 99US-0147302.  
 PR 05-AUG-1999; 99US-0147192.  
 PR 06-AUG-1999; 99US-0147260.  
 PR 06-AUG-1999; 99US-0147303.  
 PR 06-AUG-1999; 99US-0147416.  
 PR 09-AUG-1999; 99US-0147493.  
 PR 10-AUG-1999; 99US-0147935.  
 PR 11-AUG-1999; 99US-0148171.  
 PR 11-AUG-1999; 99US-0148319.  
 PR 12-AUG-1999; 99US-0148341.  
 PR 13-AUG-1999; 99US-0148565.  
 PR 13-AUG-1999; 99US-0148684.  
 PR 16-AUG-1999; 99US-0149368.  
 PR 17-AUG-1999; 99US-0149175.  
 PR 18-AUG-1999; 99US-0149426.  
 PR 20-AUG-1999; 99US-0149722.  
 PR 20-AUG-1999; 99US-0149723.  
 PR 20-AUG-1999; 99US-0149929.  
 PR 23-AUG-1999; 99US-0149902.  
 PR 23-AUG-1999; 99US-0149930.  
 PR 25-AUG-1999; 99US-0150566.  
 PR 25-AUG-1999; 99US-0150884.  
 PR 26-AUG-1999; 99US-0151065.  
 PR 27-AUG-1999; 99US-0151066.  
 PR 27-AUG-1999; 99US-0151080.  
 PR 27-AUG-1999; 99US-0151303.  
 PR 31-AUG-1999; 99US-0151438.  
 PR 01-SEP-1999; 99US-0151930.  
 PR 07-SEP-1999; 99US-0152363.  
 PR 10-SEP-1999; 99US-0153070.  
 PR 13-SEP-1999; 99US-0153758.  
 PR 15-SEP-1999; 99US-0154018.  
 PR 16-SEP-1999; 99US-0154039.

PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140891.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142820.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 21-JUL-1999; 99US-0144684.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 03-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149723.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.

PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155559.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161992.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.

Query Match 55.4%; Score 46; DB 21; Length 501;  
Best Local Similarity 68.8%; Pred. No. 41;  
Matches 11; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 3 ATSLAALVAGDAL 18  
DB 352 ATSLADGVAVAGDAI 367  
IIIIII IIII I  
I

RESULT 8  
AA642080  
ID AA642080 standard; Protein; 514 AA.  
XX  
AC AA642080;  
XX  
DT 18-OCT-2000 (first entry)  
XX  
XX Arabidopsis thaliana protein fragment SEQ ID NO: 52435.  
DE  
XX Protein identification: signal transduction pathway; metabolic pathway;  
KM hybridisation assay; genetic mapping; gene expression control; promoter;  
KM termination sequence.  
XX  
OS Arabidopsis thaliana.  
XX

PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 18-OCT-1999; 99US-0159638.  
PR 21-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161992.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.

Query Match 55 43; Score 46; DB 21; Length 443;  
Best Local Similarity 68.83; Pred. No. 36;  
Matches 11; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 3 ATSLAAALVACDAL 18  
Db 353 ATSLADGTVACDAI 368

RESULT 7  
ID AAG54408 standard; Protein; 501 AA.  
XX AAG54408:  
AC AAG54408:  
XX  
DT 18-OCT-2000 (first entry)  
XX  
DE Arabidopsis thaliana protein fragment SEQ ID NO: 69371.

XX Protein identification; signal transduction pathway; metabolic pathway;  
KW hybridisation assay; genetic mapping; gene expression control; promoter;  
KW termination sequence.  
XX Arabidopsis thaliana.  
OS Ep1033405-A2.  
PN  
XX  
PD 06-SEP-2000.  
XX 25-FEB-2000; 2000EP-0301439.  
PE  
XX 25-FEB-1999; 99US-0121825.  
PR 05-MAR-1999; 99US-0123180.  
PR 09-MAR-1999; 99US-0123548.  
PR 23-MAR-1999; 99US-0125788.  
PR 25-MAR-1999; 99US-0126264.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 28-APR-1999; 99US-0130891.  
PR 30-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 30-APR-1999; 99US-0132407.  
PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 06-MAY-1999; 99US-0132487.  
PR 07-MAY-1999; 99US-0132863.  
PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.

ID	AG65069 standard; Protein; 443 AA.	PR	18-JUN-1999;	99US-0139461;
XX		PR	18-JUN-1999;	99US-0139462;
AC	AG65069;	PR	18-JUN-1999;	99US-0139463;
XX		PR	18-JUN-1999;	99US-0139750;
DT	18-OCT-2000 (first entry)	PR	18-JUN-1999;	99US-0139763;
XX		PR	21-JUN-1999;	99US-0139817;
DE	Arabidopsis thaliana protein fragment SEQ ID NO: 70539.	PR	23-JUN-1999;	99US-0140353;
XX		PR	23-JUN-1999;	99US-0140354;
KW	Protein identification; signal transduction pathway; metabolic pathway; hybridisation assay; genetic mapping; gene expression control; promoter; termination sequence.	PR	24-JUN-1999;	99US-0140695;
KV		PR	28-JUN-1999;	99US-0140823;
XX		PR	30-JUN-1999;	99US-0141287;
OS	Arabidopsis thaliana.	PR	01-JUL-1999;	99US-0141842;
XX		PR	01-JUL-1999;	99US-0142154;
PN	EP1033405-A2.	PR	06-JUL-1999;	99US-0142055;
XX		PR	06-JUL-1999;	99US-0142390;
PD	06-SEP-2000.	PR	08-JUL-1999;	99US-0142803;
XX		PR	09-JUL-1999;	99US-0142920;
PF	25-FEB-2000; 2000EP-0301439.	PR	12-JUL-1999;	99US-0142977;
XX		PR	13-JUL-1999;	99US-0143542;
XX		PR	14-JUL-1999;	99US-0143624;
PR	05-MAR-1999;	PR	15-JUL-1999;	99US-0144005;
PR	09-MAR-1999;	PR	15-JUL-1999;	99US-0144085;
PR	23-MAR-1999;	PR	16-JUL-1999;	99US-0144325;
PR	25-MAR-1999;	PR	19-JUL-1999;	99US-0144331;
PR	01-APR-1999;	PR	19-JUL-1999;	99US-0144332;
PR	06-APR-1999;	PR	19-JUL-1999;	99US-0144333;
PR	08-APR-1999;	PR	19-JUL-1999;	99US-0144334;
PR	16-APR-1999;	PR	19-JUL-1999;	99US-0144335;
PR	19-APR-1999;	PR	20-JUL-1999;	99US-0144382;
PR	21-APR-1999;	PR	20-JUL-1999;	99US-0144814;
PR	23-APR-1999;	PR	20-JUL-1999;	99US-0145086;
PR	28-APR-1999;	PR	21-JUL-1999;	99US-0145088;
PR	30-APR-1999;	PR	21-JUL-1999;	99US-0145085;
PR	04-MAY-1999;	PR	22-JUL-1999;	99US-0145087;
PR	05-MAY-1999;	PR	22-JUL-1999;	99US-0145089;
PR	06-MAY-1999;	PR	22-JUL-1999;	99US-0145192;
PR	07-MAY-1999;	PR	23-JUL-1999;	99US-0145145;
PR	11-MAY-1999;	PR	23-JUL-1999;	99US-0145218;
PR	14-MAY-1999;	PR	26-JUL-1999;	99US-0145276;
PR	14-MAY-1999;	PR	27-JUL-1999;	99US-0145913;
PR	14-MAY-1999;	PR	27-JUL-1999;	99US-0145918;
PR	18-MAY-1999;	PR	28-JUL-1999;	99US-0145951;
PR	19-MAY-1999;	PR	02-AUG-1999;	99US-0146386;
PR	20-MAY-1999;	PR	02-AUG-1999;	99US-0146388;
PR	21-MAY-1999;	PR	03-AUG-1999;	99US-0146389;
PR	24-MAY-1999;	PR	03-AUG-1999;	99US-0147038;
PR	25-MAY-1999;	PR	04-AUG-1999;	99US-0147302;
PR	27-MAY-1999;	PR	05-AUG-1999;	99US-0147192;
PR	28-MAY-1999;	PR	06-AUG-1999;	99US-0147260;
PR	01-JUN-1999;	PR	06-AUG-1999;	99US-0147416;
PR	03-JUN-1999;	PR	09-AUG-1999;	99US-0147935;
PR	04-JUN-1999;	PR	10-AUG-1999;	99US-0148171;
PR	07-JUN-1999;	PR	11-AUG-1999;	99US-0148319;
PR	08-JUN-1999;	PR	12-AUG-1999;	99US-0148341;
PR	10-JUN-1999;	PR	13-AUG-1999;	99US-0148565;
PR	10-JUN-1999;	PR	13-AUG-1999;	99US-0148684;
PR	14-JUN-1999;	PR	16-AUG-1999;	99US-0149368;
PR	16-JUN-1999;	PR	17-AUG-1999;	99US-0149157;
PR	18-JUN-1999;	PR	18-AUG-1999;	99US-0149426;
PR	18-JUN-1999;	PR	20-AUG-1999;	99US-0149723;
PR	18-JUN-1999;	PR	23-AUG-1999;	99US-0149902;

PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139839.  
PR 23-JUN-1999; 99US-0140353.  
PR 24-JUN-1999; 99US-0140354.  
PR 28-JUN-1999; 99US-0140695.  
PR 29-JUN-1999; 99US-0140823.  
PR 30-JUN-1999; 99US-0140981.  
PR 01-JUL-1999; 99US-0141287.  
PR 02-JUL-1999; 99US-0141842.  
PR 06-JUL-1999; 99US-0142154.  
PR 06-JUL-1999; 99US-0142055.  
PR 08-JUL-1999; 99US-0142390.  
PR 09-JUL-1999; 99US-0142803.  
PR 12-JUL-1999; 99US-0142920.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 19-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144632.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 23-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 26-JUL-1999; 99US-0145224.  
PR 27-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 03-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.

PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149729.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152353.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160960.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161982.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.

## Query Match

55.4%;

Score 46; DB 21; Length 320;

Best Local Similarity 68.8%;

Pred. No. 25; Matches 11; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

OY 3 ATSLAALAVAGDAL 18

Db 158 ATSLADGVAVAGDAI 173

RESULT 6  
AAG55069

PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149930.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0158293.  
PR 13-OCT-1999; 99US-0158294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161992.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.

Query Match 55.4%; Score 46; DB 21; Length 307;  
Best Local Similarity 68.8%; Pred. No. 24;  
Matches 11; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

OY 3 ATSLAALAVAGDAL 18  
||||| |||||  
Db 158 ATSLADGVAVAGAI 173  
.....  
RESULT 5  
AAAG42081  
ID AAAG42081 standard; Protein; 320 AA.  
XX  
AC AAAG42081;  
XX  
DT 18-OCT-2000 (first entry)  
XX  
DE Arabidopsis thaliana protein fragment SEQ ID NO: 52436.  
XX  
KW Protein identification; signal transduction pathway; metabolic pathway;  
KW hybridisation assay; genetic mapping; gene expression control; promoter;  
KW termination sequence.  
XX  
OS Arabidopsis thaliana.  
XX  
PN EP1033405-A2.  
XX  
PD 06-SEP-2000.  
XX  
PF 25-FEB-2000; 2000EP-0301439.  
XX  
PE 25-FEB-1999; 99US-0121825.  
PR 05-MAR-1999; 99US-0123180.  
PR 09-MAR-1999; 99US-0123548.  
PR 23-MAR-1999; 99US-0125788.  
PR 25-MAR-1999; 99US-0126264.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 23-APR-1999; 99US-0130891.  
PR 28-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 30-APR-1999; 99US-0132407.  
PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 06-MAY-1999; 99US-0132487.  
PR 07-MAY-1999; 99US-0132863.  
PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134570.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136592.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.

PR	08-JUN-1999	9905-0138094
PR	10-JUN-1999	9905-0138540
PR	14-JUN-1999	9905-0138847
PR	14-JUN-1999	9905-0139111
PR	16-JUN-1999	9905-0139452
PR	16-JUN-1999	9905-0139493
PR	17-JUN-1999	9905-0139493
PR	18-JUN-1999	9905-0139454
PR	18-JUN-1999	9905-0139456
PR	18-JUN-1999	9905-0139456
PR	18-JUN-1999	9905-0139457
PR	18-JUN-1999	9905-0139458
PR	18-JUN-1999	9905-0139459
PR	18-JUN-1999	9905-0139460
PR	22-JUN-1999	9905-0139617
PR	23-JUN-1999	9905-0139699
PR	23-JUN-1999	9905-0140353
PR	23-JUN-1999	9905-0140354
PR	24-JUN-1999	9905-0140695
PR	28-JUN-1999	9905-0140823
PR	29-JUN-1999	9905-0140991
PR	30-JUN-1999	9905-0141287
PR	01-JUL-1999	9905-0141244
PR	01-JUL-1999	9905-0141542
PR	02-JUL-1999	9905-0142055
PR	06-JUL-1999	9905-0142390
PR	06-JUL-1999	9905-0142830
PR	09-JUL-1999	9905-0142927
PR	12-JUL-1999	9905-0143277
PR	13-JUL-1999	9905-0143542
PR	14-JUL-1999	9905-0143624
PR	15-JUL-1999	9905-0144005
PR	16-JUL-1999	9905-0144086
PR	19-JUL-1999	9905-0144325
PR	19-JUL-1999	9905-0144331
PR	19-JUL-1999	9905-0144332
PR	19-JUL-1999	9905-0144333
PR	19-JUL-1999	9905-0144334
PR	19-JUL-1999	9905-0144335
PR	20-JUL-1999	9905-0144352
PR	20-JUL-1999	9905-0144632
PR	20-JUL-1999	9905-0144684
PR	21-JUL-1999	9905-0144814
PR	21-JUL-1999	9905-0145086
PR	21-JUL-1999	9905-0145087
PR	22-JUL-1999	9905-0145087
PR	22-JUL-1999	9905-0145089
PR	22-JUL-1999	9905-0145192
PR	23-JUL-1999	9905-0145145
PR	23-JUL-1999	9905-0145218
PR	23-JUL-1999	9905-0145524
PR	26-JUL-1999	9905-0145276
PR	27-JUL-1999	9905-0145513
PR	27-JUL-1999	9905-0145519
PR	27-JUL-1999	9905-0145519
PR	28-JUL-1999	9905-0145521
PR	02-AUG-1999	9905-0146386
PR	02-AUG-1999	9905-0146388
PR	02-AUG-1999	9905-0146389
PR	03-AUG-1999	9905-0147038
PR	04-AUG-1999	9905-0147204
PR	04-AUG-1999	9905-0147302
PR	05-AUG-1999	9905-0147192
PR	05-AUG-1999	9905-0147260
PR	06-AUG-1999	9905-0147303
PR	06-AUG-1999	9905-0147416

PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139889.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144684.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145313.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149375.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 23-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150684.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 23-OCT-1999; 99US-0161404.  
PR 23-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161992.

PT  
XX  
PS Claim 9; Page 112; 185pp; English.  
XX  
CC AAF94345 to AAF94409 represent essential bacterial genes from  
CC Haemophilus influenzae, which encode the proteins given in AAB88492 to  
CC AAB88536. The present invention also describes methods for identifying  
CC essential bacterial genes (i.e. those essential to the survival of a  
CC bacterium) using a transposition system. The methods are used to  
CC identify essential genes from bacteria, especially H. influenzae (which  
CC causes otitis media, meningitis and upper respiratory tract infections)  
CC which may be used as targets for potential antimicrobial agents.  
CC AAF94410 to AAF94416 represent PCR primers used in the exemplification  
CC of the present invention.  
XX  
SQ Sequence 346 AA;  
Query Match 56.6%; Score 47; DB 22; Length 346;  
Best Local Similarity 64.7%; Pred. No. 19;  
Matches 11; Conservative 3; Mismatches 3; Indels 0; Gaps 0;  
OY 3 ATSLAALAVAGDALA 19  
:|:||||| |||:| | |  
Db 11 STALLAAGLAVSGSANA 27  
RESULT 2  
ID AAV91456  
AAV91456 standard; Protein; 346 AA.  
XX  
AC AAV91456;  
XX  
DT 18-JUN-2002 (first entry)  
XX  
DE Haemophilus influenzae essential gene #34.  
XX  
KM Essential bacterial gene; antifungal agent; antibacterial agent;  
KM antiparasitic agent; insecticidal agent; microbial infection;  
KM mucous membrane infection; otitis media; sinusitis; bronchitis;  
KM alveolitis; conjunctivitis; pneumonia; meningitis; epiglottitis;  
KM cellulitis; septic arthritis.  
XX  
OS Haemophilus influenzae.  
XX  
PN WO200218601-A2.  
XX  
PD 07-MAR-2002.  
XX  
PF 22-AUG-2001; 2001WO-US26245.  
XX  
PR 25-AUG-2000; 2000US-0649145.  
XX  
PA (ABBO ) ABBOTT LAB.  
XX  
PI Chovan LE, Hessler PE, Reich KA;  
XX  
DR WPI; 2002-304258/34.  
XX  
N-PSDB; ABR64942.  
XX  
PT Essential bacterial genes in Haemophilus influenzae necessary for  
PT bacterium's growth and survival, useful for screening inhibitors of  
PT polypeptides and developing therapeutic agents e.g. antimicrobial  
XX  
PS Claim 9; Page 112; 185pp; English.  
XX  
CC The invention describes an essential bacterial gene (I) comprising a  
CC purified polynucleotide isolated from Haemophilus influenzae where (I)  
CC is essential to H. influenzae survival. The encoded polypeptide (II) is  
CC useful for screening substances that function to inhibit essential H.  
CC influenzae polypeptides by contacting (II) with the desired substances  
CC and measuring the response by a screen from specific, enzyme, general,  
CC affinity, phenotypic and binding screen. (I), and (II) are useful in  
CC developing therapeutic agents such as antifungal, antibacterial and

CC antiparasitic agent, insecticidal agent, and preventive antimicrobial  
CC agents which are effective in preventing microbial infection or useful  
CC in treatment of that particular infection. (I) and (II) may also be  
CC useful in treatment of mucous membrane infections such as otitis media,  
CC sinusitis, bronchitis, alveolitis, conjunctivitis, pneumonia, meningitis,  
CC epiglottitis, cellulitis and septic arthritis. This is the amino acid  
CC sequence of an essential H. influenzae gene, described in the invention.  
XX  
SQ Sequence 346 AA;  
Query Match 56.6%; Score 47; DB 23; Length 346;  
Best Local Similarity 64.7%; Pred. No. 19;  
Matches 11; Conservative 3; Mismatches 3; Indels 0; Gaps 0;  
OY 3 ATSLAALAVAGDALA 19  
:|:||||| |||:| | |  
Db 11 STALLAAGLAVSGSANA 27  
RESULT 3  
ID AAG54410  
AAG54410 standard; Protein; 299 AA.  
XX  
AC AAG54410;  
XX  
DT 18-OCT-2000 (first entry)  
XX  
DE Arabidopsis thaliana protein fragment SEQ ID NO: 69373.  
XX  
KM Protein identification; signal transduction pathway; metabolic pathway;  
KM hybridisation assay; genetic mapping; gene expression control; promoter;  
KM termination sequence.  
XX  
OS Arabidopsis thaliana.  
XX  
PN EP1033405-A2.  
XX  
PD 06-SEP-2000.  
XX  
PF 25-FEB-2000; 2000EP-0301439.  
XX  
PR 25-FEB-1999; 99US-0121825.  
PR 05-MAR-1999; 99US-0123180.  
PR 09-MAR-1999; 99US-0123548.  
PR 23-MAR-1999; 99US-0125788.  
PR 25-MAR-1999; 99US-0126264.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 23-APR-1999; 99US-0130891.  
PR 28-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 30-APR-1999; 99US-0132407.  
PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 06-MAY-1999; 99US-0132487.  
PR 07-MAY-1999; 99US-0132863.  
PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.

GenCore version 5.1.4-p5.4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 9, 2003, 15:10:20 ; Search time 2.81963 Seconds

(without alignments)  
897.904 Million cell updates/sec

Title: US-10-028-245-3

Perfect score: 83

Sequence: 1 MRATSLAALAVAGDALA 19

Scoring table: BLOSUM62

Gapop 10.0, Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A.Geneseq\_101002:\*

1: /SID2/gcgdata/geneeq/geneeqp-emb1/AA1980.DAT:\*  
2: /SID2/gcgdata/geneeq/geneeqp-emb1/AA1981.DAT:\*  
3: /SID2/gcgdata/geneeq/geneeqp-emb1/AA1982.DAT:\*  
4: /SID2/gcgdata/geneeq/geneeqp-emb1/AA1983.DAT:\*  
5: /SID2/gcgdata/geneeq/geneeqp-emb1/AA1984.DAT:\*  
6: /SID2/gcgdata/geneeq/geneeqp-emb1/AA1985.DAT:\*  
7: /SID2/gcgdata/geneeq/geneeqp-emb1/AA1986.DAT:\*  
8: /SID2/gcgdata/geneeq/geneeqp-emb1/AA1987.DAT:\*  
9: /SID2/gcgdata/geneeq/geneeqp-emb1/AA1988.DAT:\*  
10: /SID2/gcgdata/geneeq/geneeqp-emb1/AA1989.DAT:\*  
11: /SID2/gcgdata/geneeq/geneeqp-emb1/AA1990.DAT:\*  
12: /SID2/gcgdata/geneeq/geneeqp-emb1/AA1991.DAT:\*  
13: /SID2/gcgdata/geneeq/geneeqp-emb1/AA1992.DAT:\*  
14: /SID2/gcgdata/geneeq/geneeqp-emb1/AA1993.DAT:\*  
15: /SID2/gcgdata/geneeq/geneeqp-emb1/AA1994.DAT:\*  
16: /SID2/gcgdata/geneeq/geneeqp-emb1/AA1995.DAT:\*  
17: /SID2/gcgdata/geneeq/geneeqp-emb1/AA1996.DAT:\*  
18: /SID2/gcgdata/geneeq/geneeqp-emb1/AA1997.DAT:\*  
19: /SID2/gcgdata/geneeq/geneeqp-emb1/AA1998.DAT:\*  
20: /SID2/gcgdata/geneeq/geneeqp-emb1/AA1999.DAT:\*  
21: /SID2/gcgdata/geneeq/geneeqp-emb1/AA2000.DAT:\*  
22: /SID2/gcgdata/geneeq/geneeqp-emb1/AA2001.DAT:\*  
23: /SID2/gcgdata/geneeq/geneeqp-emb1/AA2002.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	47	56.6	346	22 AAB88525	Haemophilus influenzae
2	47	56.6	346	23 AAU91456	Haemophilus influenzae
3	46	55.4	299	21 AAG54410	Arbidopis thaila
4	46	55.4	307	21 AAG54409	Arbidopis thaila
5	46	55.4	320	21 AAG42081	Arbidopis thaila
6	46	55.4	443	21 AAG55059	Arbidopis thaila
7	46	55.4	501	21 AAG54408	Arbidopis thaila
8	46	55.4	514	21 AAG42080	Arbidopis thaila
9	46	55.4	515	21 AAG42079	Arbidopis thaila
10	45	54.2	166	23 ABB54458	Lactococcus lactis

11	44	53.0	271	23 AAG68270	Human POLY17 prote
12	44	53.0	578	22 AAB66692	Beta-fructofuranos
13	43	51.8	98	19 AAW76094	Human osteocalcin
14	43	51.8	100	23 AAU10687	Reference sequence
15	43	51.8	100	23 AAU10688	Polyomphic varian
16	43	51.8	137	21 AAG43076	Arbidopis thaila
17	43	51.8	226	12 AAR13466	Klebsiella pneumoniae
18	43	51.8	272	12 AAR13466	Prohibitin, Rattus
19	43	51.8	272	14 AAR42215	Human prohibitin
20	43	51.8	279	21 AAB43874	Human cancer assoc
21	43	51.8	279	22 AAG73845	Human cancer assoc
22	43	51.8	284	22 AEG04798	Novel human diago
23	43	51.8	294	22 AAU38215	Salmonella typhi c
24	43	51.8	297	22 AAU30413	Novel human secret
25	43	51.8	343	22 AAG18633	Novel human diago
26	43	51.8	471	19 AAW54352	Heat shock 27 kD p
27	42	50.6	25	22 AAB81158	SLPA peptide from
28	42	50.6	80	21 AAG57280	Arbidopis thaila
29	42	50.6	137	21 AAG1637	Arbidopis thaila
30	42	50.6	358	19 AAW56155	Arbidopis thaila
31	42	50.6	712	22 AAM57843	Human bone marrow
32	42	50.6	712	22 AAM70260	Human bone marrow
33	42	50.6	712	22 AAM18088	Peptide #4522 enco
34	42	50.6	712	23 AEG39902	Human peptide enco
35	42	50.6	891	23 AAU99614	Human glioma anti
36	42	50.6	1204	22 AAM78475	Human protein SEQ
37	41.5	50.0	70	19 AAM61628	Clone HROAD63 of L
38	41.5	50.0	128	17 AAM06292	Human squamous cel
39	41	49.4	346	22 AAU36390	Pseudomonas aerugi
40	41	49.4	563	21 AAB10080	F. bidentis glutam
41	41	49.4	617	21 AAG28574	Arbidopis thaila
42	41	49.4	629	22 AAB66935	Human GLUTX2, Hom
43	41	49.4	629	22 AAB66940	GLUTX2 consensus s
44	41	49.4	648	23 AAG61548	Human transporter
45	41	49.4	762	21 AAG28573	Arbidopis thaila

#### ALIGNMENTS

RESULT 1  
AAB88525  
ID AAB88525 standard; Protein: 346 AA.  
AC AAB88525;  
DT 04-JUN-2001 (first entry)  
DE Haemophilus influenzae essential bacterial protein SEQ ID NO:68.  
XX  
DE Haemophilus influenzae; essential bacterial gene; identification:  
XX  
KW ctitis media; meningitis; upper respiratory tract infection;  
KW infection; antimicrobial.  
XX  
XX  
OS Haemophilus influenzae.  
XX  
PN WO200111033-A2.  
PD 15-FEB-2001.  
PF 03-AUG-2000; 2000WO-US21176.  
PR 04-AUG-1999; 99US-0368382.  
XX  
XX (ABBO ) ABBOTT LAB.  
XX Chovan LE, Hessler PE, Reich KA;  
XX WPI: 2001-147511/15.  
XX N-PSDB: AAF94378.  
XX  
XX Essential bacterial genes from Haemophilus influenzae and methods for  
XX identifying 'essential' genes that may be potential therapeutic targets  
XX

THIS PAGE BLANK (USPTO)

```

Db 190 SOYIEFGNSWTCGAWTW--TNVNDNMKSLTDP---SDKIIEHNOYLDSDSGSATSVCVS 244
OY 232 DNV--DAFNDFADMLRQNRKRAIISSETGASMESSCMATFACQAKAISENDVYIGFVGWG 289
Db 245 STIGERITSANTQMLNKGKGIIGEPAGAGANDVCETAITGMIDYMAQNTDVTGAIWMA 304
OY 290 AGSFDTSYLLTLPPLGKPGNVYTNKMLNECILDQFT 325
Db 305 AGPMWGDYIFSMER-----DNGIAYQOILPLLT 332

```

## RESULT 14

```

OQ9P81 PRELIMINARY; PRT; 385 AA.
AC OQ9P81;
DT 01-OCT-2000 (TREMBlrel. 15, Created)
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
DE 01-MAR-2002 (TREMBlrel. 20, Last annotation update)
DE Endoglucanase 1.
OS Robillarda sp. (strain Y-20).
OC Eukaryota; Fungi; Ascomycota; mitosporic Ascomycota; Robillarda.
OX NCBI_TaxId=72589;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Y-20;
RA Kashiwagi Y.;
RT "Endoglucanase gene from cellulolytic fungi, Robillarda sp. Y-20.";
RL Submitted (AUG-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB030819; BAA90480.1; -.
DR InterPro: IPR001547; GH_5.
DR Pfam: PF00150; cellulase; 1.
DR PROSITE: PS00659; GLYCOSYL_HYDROL_F5; UNKNOWN.1.
SQ SEQUENCE 385 AA; 41428 MW; 149604D42369AD33 CRC64;

```

Query Match 15.6%; Score 346; DB 3; Length 385;  
 Best Local Similarity 28.5%; Pred. No. 2.5e-16;  
 Matches 99; Conservative 62; Mismatches 134; Indels 52; Gaps 13;

```

OY 3 IKVLGVAIRGIDFGCDIDGSCPTDTSSVPLSTYKGGDGA----GQKKHFAEDDGLNVERI 58
Db 19 IYAGVAESSGEGV----WSATQTPCTGLPGFEGVDAFISAAVDVHYDQHLNLEFV 74
OY 59 SATWQFVLNNT-----VDGKLDELNMGSYNKVYNACLET--GAYCMIDMHNFRYRN----- 107
Db 75 A-----FLERMCPRATGLAAGFNEHFDYFKENADVITYTKGAYALIDPRNRYNDPSY 130
OY 108 ----GGIIG----QGVSDDIFFVDLWQIAKKYEDNDKIIIFGLMNEPHDIDIEIWAOTCQ 159
Db 131 OPGSGSVICNTSDSTAATTEQFGEFEGELASRFENDNERVYFGLMNEPHDMATSLVLANQ 190
OY 160 KVTATIRKKGATSQMILLPGTNFASVETV--STGSAAELAGKITNDGSTDLLXPDVHKY 217
Db 191 AADDAIRANA-SNLTIIMPNSWTHGSHWTEGSDPSSALLNQRPDLNNTAI--DIHEY 246
OY 218 LDINNSGSHAECTTDVDAFNDFADMLRQNRKRAIISSETGASMEPSCMTAFCAQNKAISE 277
Db 247 LDYDFSGHLECYSDPETNLAALTALAKENNLKAFITTEFGSNTSCQEMLPOLINMAD 306
OY 278 NSDVYIGVVGWAGSF-----DTSYTLTLTP-----LGRPCNV 310
Db 307 NAE-YIGMTAMAAAGPFMGPSCTNSTQSLSEPGSTAVDGSPLGLY 352

```

## RESULT 15

```

OQ9Y86 PRELIMINARY; PRT; 325 AA.
AC OQ9Y86;
DT 01-NOV-1999 (TREMBlrel. 12, Created)
DT 01-NOV-1999 (TREMBlrel. 12, Last sequence update)
DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)
DE Endo-beta-1,4-glucanase A precursor.

```

```

GN BGIA.
OS Emericella nidulans (Aspergillus nidulans).
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
OC Eurotiales; Trichocomaceae; Emericella.
OX NCBI_TaxId=5072;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99314152; PubMed=10386374;
RA Chikamatsu G., Shirai K., Kato M., Kobayashi T., Tsukagoshi N.;
RT "Structure and expression properties of the endo-beta-1,4-glucanase A gene from the filamentous fungus Aspergillus nidulans.";
RL FEWS Microbiol. Lett. 175:239-245(1999).
DR EMBL; AB009402; BAA82592.1; -.
DR InterPro: IPR001547; GH_5.
DR Pfam: PF00150; cellulase; 1.
KM Signal.
FT SIGNAL. 1 19 POTENTIAL.
SQ SEQUENCE 325 AA; 35648 MW; C5582A85DAA26829 CRC64;

```

Query Match 15.6%; Score 345.5; DB 3; Length 325;  
 Best Local Similarity 27.2%; Pred. No. 2.2e-16;  
 Matches 83; Conservative 65; Mismatches 136; Indels 21; Gaps 7;

```

OY 1 GKIKYLGVAIRGIDFGCDIDGSCPTDTSSVPLSTYKGGDGAQKKHFAEDDGLNVERISA 60
Db 19 GAFTWLGTNEAGAEFG--EGSYRGEIGTEYIWPDLGTICT-----LRREGMNIFFVAF 69
OY 61 TNPQFVLNNTVDGKLDELNMGSYNKVYNACLETGAYCMIDMHNFRYNGIIGGQVSDI 120
Db 70 SMERLVDPDSLAVGVADEYFQDLYETVNGITLALGAVYVLDPHNNGRY-GNTI---TSTD 124
OY 121 FVDLWQIAKKYEDNDKIIIFGLMNEPHDIDIEIWAOTCQKVTATIRKKGATSQMILLPGT 180
Db 125 FAFWTLTAEFASNELVIFDTNNEYHTMQSLVNLNQALDAIRASGATSGYIFAEQN 184
OY 181 NFASVETVYSTGSAEALGKITNDGSTDLLYFVHKYLDINNSGSHAECTTDV--DAFN 238
Db 185 SWTGAMTWVDVN--DNMKALITDPQ---DKLIYEMHQYLDSDSGSGTWTACVSTIGSERVT 239
OY 239 DFADMLRQNRKRAIISSETGASMEPSCMTAFCAQNKAISENDVYIGFVGAGASPTSYI 298
Db 240 AATNMLRENGKLGVLDEFGAGANNQCKDAVADLLEYLENSDVLGALMWAAGPMMGDY 299
OY 299 LTLTP 303
Db 300 FNMER 304

```

Search completed: May 9, 2003, 15:20:48  
 Job time : 105.358 secs

OX NCBI\_TaxID=5061;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=MO 94/14953;  
 RX MEDLINE=98432774; PubMed=9758775;  
 RA van Peijl N.N.M.E., Gielkens M.M.C., de Vries R.P., Visser J.,  
 De Graaf L.H.,  
 RT "The transcriptional activator XlnR regulates both xylenolytic and  
 endoglucanase gene expression in *Aspergillus niger*."  
 RL Appl. Environ. Microbiol. 64:3615-3619(1998).  
 CC -1- FUNCTION: HAS CARBOXYLMEHYDROLYTIC ACTIVITY.  
 CC -1- CATALYTIC ACTIVITY: ENDOMETHYLOLYSIS OF 1,4-BETA-D-GLUCOSIDIC  
 CC LINKAGES IN CELLULOSE.  
 CC -1- SIMILARITY: BELONGS TO CELLULOSE FAMILY H (FAMILY 12 OF GLYCOSYL  
 CC HYDROLASES).  
 DR EMBL: AJ224452; CA11965.1;  
 DR InterPro: IPR001547; GH\_5.  
 DR Pfam: PF00150; cellulase; 1.  
 KM Cellulose degradation; Hydrolyase; Glycosidase; Signal.  
 FT SIGNAL  
 FT CHAIN 1 18 POTENTIAL.  
 FT SEQUENCE 19 331 ENDOLUCANASE B.1  
 SQ SEQUENCE 331 AA; 36558 MW; F3AE4BF70007C707 CRC64;  
 Query Match 16.38; Score 361; DB 3; Length 331;  
 Best Local Similarity 29.08; Pred. No. 1.0e-17;  
 Matches 89; Conservative 56; Mismatches 130; Indels 32; Gaps 7;  
 OY 4 KYLVVAIPGIDFGCDIGSCPT-----DTSVPLISYKGGDAGQKHFADDDGLNVERI 58  
 DB 33 EMFSSNSGAEFGTNGICWGTDTYIFPDPTIITLICK-----GKNFPRV 77  
 OY 59 SATWQFVLANITVDKLDLNLNWSYKVNACLETGAYCMIDHNFARYNGIIGQGVSD 118  
 DB 78 QFMNERLLPDSMTGSYDEEYLANLTYYKAVTDGGAHALIDPHNYGRNGEIIIS--ST 133  
 OY 119 DIFDVLWQIAKYEDNDKIIIFGLMNEPHDIEIHAOTCQKVYTAIRKACATSCMILLP 178  
 DB 134 SDFQTFMQLNLAGQYKNDLVFDTNNEYXMDODLVNLNQAALINGRAAGA-SQYIFVE 192  
 OY 179 GTNFASVETVSTGSABALGKITNPDGSTDLLYFDVHKYLDINNSGSHAECTDNV--DA 236  
 DB 193 GNSMTGAMTWVDN--DNMKMLTDPD---DKIYVEMHOYLDSDSGTSRTSETVSGTIGER 247  
 OY 237 FNDPADMLRQNRQAIISSETGASMEPSQMTAFCAONKAISENSDVIYIGFVGWAGSEPTD 296  
 DB 248 ITDATQMLKNDKKVGTIGEYAGSNDYCRSAVSQMLEYMANNTDVMKASWMAAGPMMGD 307  
 OY 297 YITLTP 303  
 DB 308 YITSLEP 314  
 RESULT 12  
 O59951 PRELIMINARY; PRT; 332 AA.  
 AC O59951;  
 DT 01-AUG-1998 (TREMBLrel. 07, Created)  
 DT 01-AUG-1998 (TREMBLrel. 07, Last sequence update)  
 DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)  
 DE Endoglucanase V (EC 3.2.1.4).  
 GN ENGL.  
 OS *Aspergillus aculeatus*.  
 OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;  
 OC Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; *Aspergillus*.  
 OX NCBI\_TaxID=5053;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=MO 94/14953;  
 RA Kofod L.V., Dalboge H., Andersen L.N., Kauppinen S., Christgau S.,  
 RT *Aspergillus aculeatus* endoglucanase."  
 RL Submitted (APR-1998) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.

RC STRAIN=MO 94/14953;  
 RA Muller S., Sandal T., Kamp-Hansen P., Dalboge H.,  
 RL Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AF054512; AAC08587.1;  
 DR InterPro: IPR001547; GH\_5.  
 DR Pfam: PF00150; cellulase; 1.  
 KM Glycosidase; Hydrolyase.  
 FT SEQUENCE 332 AA; 36188 MW; 45A0CABBE5B3AD21 CRC64;  
 Query Match 16.08; Score 354; DB 3; Length 332;  
 Best Local Similarity 29.08; Pred. No. 5.7e-17;  
 Matches 88; Conservative 53; Mismatches 132; Indels 30; Gaps 7;  
 OY 7 GVAIPGIDFGCDIGSCPTDTSVPLISYKGGDAGQKHFADDDGLNVERIATWQFVL 66  
 DB 44 GTALPDT---WGDIYIF-DTSALATVSK-----GNIIFVQFMNELV 84  
 OY 67 NNTYDGLKLDLNLNWSYKVNACLETGAYCMIDHNFARYNGIIGQGVSDIIFDLNV 126  
 DB 85 PMSMTGSYDADYVLANLTYYVNAIAAGVHAIVDPHNYGRNGEIIIS--STADFQTFWQ 140  
 OY 127 QIAKYEDNDKIIIFGLMNEPHDIEIHAOTCQKVYTAIRKAGATSCMILLPNTFAVE 186  
 DB 141 NLAGQFKNDDLVIFDTNNEYNTMQFTLVLDLNLQAALIDGIRAGATSOYIFAEQSWSGAW 200  
 OY 187 TYVSTGSABALGKITNPDGSTDLLYFDVHKYLDINNSGSHAECTDNVDA--FNDPADML 244  
 DB 201 TWADLN--DNMKMLTDPD---DKIYVEMHOYLDSDSGTSQVSETIGARLQAATQWL 255  
 OY 245 RQNRQAIISSETGASMEPSQMTAFCAONKAISENSDVIYIGFVGWAGSEPTSYITLTP 304  
 DB 256 KDNQGVLDILGEYAGANDVCRTALAGMLEYMANNTDVMKASWMAAGPMMGDYMSMP 315  
 OY 305 GKP 307  
 DB 316 SGP 318  
 RESULT 13  
 O87G26 PRELIMINARY; PRT; 335 AA.  
 AC O87G26;  
 DT 01-JUN-2002 (TREMBLrel. 21, Created)  
 DT 01-JUN-2002 (TREMBLrel. 21, Last sequence update)  
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)  
 DE Endo type cellulase ENGL.  
 GN ENGL.  
 OS *Thermascus aurantiacus*.  
 OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;  
 OC Eurotiales; Trichocomaceae; *Thermascus*.  
 OX NCBI\_TaxID=5087;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Hong J., Tanaki H., Yamamoto K., Kumagai H.,  
 RT "Endo-beta-1,4-glucanase genomic DNA from *Thermascus aurantiacus*  
 RT IF09748."  
 RL Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AF487830; AAL88714.1;  
 SQ SEQUENCE 335 AA; 37017 MW; 907217B06BA7C3D CRC64;  
 Query Match 15.88; Score 349.5; DB 3; Length 335;  
 Best Local Similarity 29.38; Pred. No. 1.2e-16;  
 Matches 81; Conservative 50; Mismatches 126; Indels 19; Gaps 5;  
 OY 52 GLNFRISATWQFVLANITVDKLDLNLNWSYKVNACLETGAYCMIDHNFARYNGI 111  
 DB 74 GNIIFRVPPMERLVPMSMTGSPDPNTLADITATVNAITQGAAYVDPHYGTYVSTI 133  
 OY 112 GQGVSDIIFDVLWQIAKYEDNDKIIIFGLMNEPHDIEIHAOTCQKVYTAIRKAGAT 171  
 DB 134 S-----SPSDFQTFWKYTAASQFASPLVIFDTNNEYHMDQTLVNLNQAALIDGIRSGAT 189  
 OY 172 SOMILLPNTFAVEYVSTGSABALGKITNPDGSTDLLYFDVHKYLDINNSGSHAECT 231

OY 190 STGSALGKRTNPDGSTDLYEDVHKYIDINNSGSHAECTDNV--DAFNEDFADWLKRON 247  
 DB 207 QVN--DAMNLTDPQNK---IYEMHQYLDSDSGSTSDCVNNTIGQDRVASTAMUKON 261  
 OY 248 KROAITSEFGASMEPSCMTAFCAONKAISENSVYIGFVGWAGSEFDTSYLTLP 303  
 DB 262 GKRAIIGEFAGANSVCSAVTGLDLHADNTDVTGAIWMAGPMMVANYIFSMDP 317

## RESULT 9

OY 99C328 PRELIMINARY: PRT: 389 AA.  
 AC 99C328;  
 DT 01-JUN-2001 (Tremblrel. 17, Created)  
 DT 01-JUN-2001 (Tremblrel. 17, Last sequence update)  
 DE 01-JUN-2002 (Tremblrel. 21, Last annotation update)  
 OS Endoglucanase (EC 3.2.1.4).  
 OC Volvariella volucae.  
 OC Eukaryota; Fungi; Basidiomycota; Hymenomycetes; Homobasidiomycetes;  
 OC Agaricales; Pluteaceae; Volvariella.  
 OX NCBI\_Taxid=36659;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-V14;  
 RA Ding S., Ge W., Buswell J.A.;  
 RT "Purification of endoglucanase, and molecular cloning and expression  
 of an endoglucanase gene, from the edible straw mushroom, Volvariella  
 volucae."  
 RL Submitted (DEC-2000) to the EMBL/Genbank/DBJ databases.  
 DR EMBL; AF328732; AAG59832.1;  
 DR HSP; P00725; I426.  
 DR InterPro: IPR000254; CBD\_fungal.  
 DR Pfam: PF00734; CBM.1; 1.  
 DR Pfam: PF00150; cellulase; 1.  
 DR ProDom: PD001821; CBD\_fungal; 1.  
 DR SMART; SM00236; fCBM; 1.  
 DR PROSITE; PS00562; CBD\_FUNGAL; 1.  
 DR PROSITE; PS00659; GLYCOSYL\_HYDROL\_F5; UNKNOWN; 1.  
 DR GLYcosidase; Hydrolase.  
 KW SEQUENCE 389 AA; 41795 MW; 4B1B801829FE5408 CRC64;

Query Match 16.8%; Score 371.5; DB 3; Length 389;  
 Best Local Similarity 27.3%; Pred. No. 4.3e-18;  
 Matches 94; Conservative 62; Mismatches 125; Indels 63; Gaps 10;

OY 21 GSCPTTSSVPLSTYKGGGAGOMKHF-----AE----- 49  
 DB 61 GAGPTTSSAPNPTSSGCPNATKFRFGVNOAGAEFGENYIPGELGHTYWPSPSIDYF 120  
 OY 50 -DDGLVFRISATWQFVLNNT-----VDGKIDELMNGSYNKVYNACLETGAYCMIDMHN 102  
 DB 121 VNGGFYTFVRA---FKIERLSPGTLGPGFQATLNGKLTIVNTITGKNATAVADPHN 176  
 OY 103 FARNGGIIGGGVSDIDFVLDVVOIAKXYEDNDKIIFGLMNEPHDLEIMAQTCOKV 162  
 DB 177 YMYNNGNVI---TSTNSFQTMWNKLATERFSRNTVIEDVMEPYIDASVYNLNOAI 232  
 OY 163 TAIRKGAISOMILLPGTNFASVETVYSGSALGKITNPDGSTDLYEDVHKYIDINN 222  
 DB 233 NGIRAGATSQLIVGCTAMTGAMSWESSGNGAVFGAIRDPNNNTAI---EMHQYLDSDS 289  
 OY 223 SCSHAC--TTDNVDAFNPADMLRONKROAISETGASMEPSCMT---APCAONKAIS 276  
 DB 290 SGTSAICVSTYGERLRTATDMLRRNNLKGFIGEWAGSNDVCAIAVKGALCAM----- 344  
 OY 277 ENSDYVIGFVGWAGSEFDTSYLTLPKPGNYTNDKLNKECI 320  
 DB 345 QDSGWIGYGLWMAAGPMMGTIYFOSIEP---PNCASITARIILPEAL 385

RESULT 10  
 074169

ID 074169 PRELIMINARY: PRT: 412 AA.  
 AC 074169;  
 DT 01-NOV-1998 (Tremblrel. 08, Created)  
 DT 01-NOV-1998 (Tremblrel. 08, Last sequence update)  
 DE 01-JUN-2002 (Tremblrel. 21, Last annotation update)  
 OS Fli-CMCase.  
 GN CMC2.  
 OS Aspergillus aculeatus.  
 OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;  
 OC Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.  
 OX NCBI\_Taxid=5053;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Aral M., Takada G., Kwaguchi T., Sumitani J.;  
 RT "Fli-CMCase gene from Aspergillus aculeatus";  
 RL Submitted (JUN-1998) to the EMBL/Genbank/DBJ databases.  
 DR EMBL; AB015510; BAA29030.1;  
 DR HSP; P00725; 2CBH.  
 DR InterPro: IPR000254; CBD\_fungal.  
 DR InterPro: IPR001547; GH\_5.  
 DR Pfam: PF00734; CBM.1; 1.  
 DR Pfam: PF00150; cellulase; 1.  
 DR ProDom: PD001821; CBD\_fungal; 1.  
 DR SMART; SM00236; fCBM; 1.  
 DR PROSITE; PS00562; CBD\_FUNGAL; 1.  
 KW SEQUENCE 412 AA; 44128 MW; F964A5FAB44AE5EA CRC64;

Query Match 16.8%; Score 366.5; DB 3; Length 412;  
 Best Local Similarity 27.4%; Pred. No. 1e-17;  
 Matches 108; Conservative 63; Mismatches 164; Indels 59; Gaps 12;

OY 5 YLVAIRPGIDFG--CDIDGSCPT-----DTSSVPLSTYKGGGAGOMKHFADDDGLANFRI 58  
 DB 27 WLSNESGAEFGQANIPGVLTIDYMPDASKIQVL-----VNGMNTFRV 71  
 OY 59 SATWQFVLNNTVDGKIDELMNGSYNKVYNACLETGAYCMIDMHNFARNGGIIGGGVSD 118  
 DB 72 PFMERLIPTSLGADATYLAOLKATIESITSLGAVAVDPNHFGRYGNII---TST 127  
 OY 119 DIVVDLMVQ-----AKYEDNDKIIFGLMNEPHDLEIMAQTCOKVYATRKAGATS 172  
 DB 128 SDFAAFWYTYVSCPSHHYTHKDH---ADNEYHDDQTLVNLNOAATAIRAGATS 183  
 OY 173 QMILLPCTNFASVETVYSGSALGKITNPDGSTDLYEDVHKYIDINNSGSHAECTTD 232  
 DB 184 QYIFVEBNSYSGAMWTYTF--NLVNLTP---SDKIIEYEMHQYLDSDSGSTSDCVST 238  
 OY 233 NVDAFNDEA--DWLRONKROAISETGASMEPSCMTAFCAONKAISENSVYIGFVGWGA 290  
 DB 239 TIGAERVAATWLRONKRAVGEFAGANSOCLTAVTGMLDALSAASDVWIGALMWSA 298  
 OY 291 GSFDTSYLTLP--LGRPGYTDNKLMECILDFOITLDEKRYRPPISITRAEETATATA 349  
 DB 299 GPWWGYIIFNEPPTGTAVTY-----YLSLOSIVPGSSGT---TTTAVA 342  
 OY 350 TSDGDA PSTTKPIFREBTASPTPNAYTKPSDPT 383  
 DB 343 TTTTSKASTSTTTKTTTSTTTTATSTSTA 376

RESULT 11  
 074706 PRELIMINARY: PRT: 331 AA.

AC 074706;  
 DT 01-NOV-1998 (Tremblrel. 08, Created)  
 DT 01-NOV-1998 (Tremblrel. 08, Last sequence update)  
 DE 01-JUN-2002 (Tremblrel. 20, Last annotation update)  
 OS Endoglucanase B precursor (EC 3.2.1.4) (CELLULASE).  
 DE (Endo-1,4-beta-glucanase B) (CARBOXYMETHYLCELLULASE).  
 GN BGLB.  
 OS Aspergillus niger.  
 OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;  
 OC Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.

OC Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.  
 OX NCBI\_TaxID=5061;  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-IF031125;  
 RA Jiong H., Tamaki H., Akiba S., Yamamoto K., Kumagai H.;  
 RT "cDNA sequence of a thermal-stable, protease resistant  
 endoglucanase."  
 RL Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AF31518; AAG50051.1;  
 DR InterPro: IPR001547; GH\_5.  
 DR Pfam: PF00150; cellulase; 1.  
 KW Signal.  
 FT SIGNAL  
 SQ SEQUENCE 332 AA; 36700 MW; 487246561851A0D CRC64;

Query Match 17.2%; Score 381.5; DB 3; Length 332;  
 Best Local Similarity 28.7%; Pred. No. 6.9e-19;  
 Matches 88; Conservative 59; Mismatches 129; Indels 31; Gaps 6;

QY 4 KYLGVAIPGIDFGCDIDGSCPT-----DTSSVPLLSYKGGDAGQMKHFAEDDGLNFERI 58  
 DB 33 EMFGSNESGAEFGTINIGVWCTDIYIFPDPSAISTLI-----DKGMNFERV 77  
 QY 59 SATQOFLVNTVQCKLDELNMGSYNKVNACLETGATCMIDMNFARNGIIGCGVSD 118  
 DB 78 QFMERLLPDSMTGSYDEEYLANLTIVYKAVTDGAHALVDPHNYGRNGEILIS-----ST 133  
 QY 119 DIFVDLWVQIAKYEDNDKIIIFGLMNEPHDIEIMAQTCOKVYTAIRKAGATSOIILLP 178  
 DB 134 SDPOTFEMENLAGQYKNDLVAFDTNNEYHMDQDLVNLNOAALNGIRAGATSOIYFVE 193  
 QY 179 GTNFAVETVYSTGSAAALGKTNPDSSTDLVDFVHKYLDINNNGSHAECTTDNV--DA 236  
 DB 194 GNSWTGAMTWVDVN--DNMKMLTDPD--DKIYEMHQYLDSDSGSETCVSETICKER 248  
 QY 237 FNDPADMLRQNKRAIISFGASMEPSCMTAFCAQNKRAISNSVYIGFVWGAGSDTS 296  
 DB 249 VTEATQWLKDKKVGFIQYAGGSNDYCRSAVSGMLEYMANNTDVMKASWMAAGPWWGD 308  
 QY 297 YILFLTP 303  
 DB 309 YIFSLP 315

RESULT 7  
 096W08 PRELIMINARY; PRT; 332 AA;  
 AC 096W08;  
 DT 01-DEC-2001 (TREMBlrel. 19, Created)  
 DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)  
 DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)  
 DE Endoglucanase C.  
 GN ECLC  
 OS Aspergillus kawachii.  
 OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;  
 OC Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.  
 OX NCBI\_TaxID=40384;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Ito K., Shimoi H., Iwashita K., Hara Y., Hinoki Y.;  
 RT "Cloning and characterization of endoglucanase genes from an  
 industrial fungus Aspergillus kawachii."  
 RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AB055433; BAB62319.1;  
 DR InterPro: IPR001547; GH\_5.  
 DR Pfam: PF00150; cellulase; 1.  
 DR SQUENCE 332 AA; 36733 MW; A709672P5DEB1830 CRC64;

Query Match 17.1%; Score 379.5; DB 3; Length 332;  
 Best Local Similarity 28.3%; Pred. No. 9.5e-19;  
 Matches 87; Conservative 60; Mismatches 129; Indels 31; Gaps 6;

QY 4 KYLGVAIPGIDFGCDIDGSCPT-----DTSSVPLLSYKGGDAGQMKHFAEDDGLNFERI 58  
 DB 33 EMFGSNESGAEFGTINIGVWCTDIYIFPDPSAISTLI-----DKGMNFERV 77  
 QY 59 SATQOFLVNTVQCKLDELNMGSYNKVNACLETGATCMIDMNFARNGIIGCGVSD 118  
 DB 78 QFMERLLPDSMTGSYDEEYLANLTIVYKAVTDGAHALVDPHNYGRNGEILIS-----ST 133  
 QY 119 DIFVDLWVQIAKYEDNDKIIIFGLMNEPHDIEIMAQTCOKVYTAIRKAGATSOIILLP 178  
 DB 134 SDPOTFEMENLAGQYKNDLVAFDTNNEYHMDQDLVNLNOAALNGIRAGATSOIYFVE 193  
 QY 179 GTNFAVETVYSTGSAAALGKTNPDSSTDLVDFVHKYLDINNNGSHAECTTDNV--DA 236  
 DB 194 GNSWTGAMTWVDVN--DNMKMLTDPD--DKIYEMHQYLDSDSGSETCVSETICKER 248  
 QY 237 FNDPADMLRQNKRAIISFGASMEPSCMTAFCAQNKRAISNSVYIGFVWGAGSDTS 296  
 DB 249 VTEATQWLKDKKVGFIQYAGGSNDYCRSAVSGMLEYMANNTDVMKASWMAAGPWWGD 308  
 QY 297 YILFLTP 303  
 DB 309 YIFSLP 315

RESULT 8  
 08WZD7 PRELIMINARY; PRT; 334 AA.  
 ID 08WZD7;  
 AC 08WZD7;  
 DT 01-MAR-2002 (TREMBlrel. 20, Created)  
 DT 01-MAR-2002 (TREMBlrel. 20, Last sequence update)  
 DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)  
 DE Endoglucanase (EC 3.2.1.4) (Endo beta 1,4 glucanase).  
 GN EGI.  
 OS Talaromyces emersonii.  
 OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;  
 OC Eurotiales; Trichocomaceae; Talaromyces.  
 OX NCBI\_TaxID=68825;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Murray P.G., Collins C.M., Tuohy M.G.;  
 RT "Molecular cloning and expression of endoglucanase genes of  
 Talaromyces emersonii."  
 RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA Murray P.G., Collins C.M., Tuohy M.G.;  
 RT "Molecular cloning of the cellulase degrading system of Talaromyces  
 emersonii."  
 RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AF440003; AAL33630.1;  
 DR EMBL: AF440013; AAL33639.1;  
 DR InterPro: IPR001547; GH\_5.  
 DR Pfam: PF00150; cellulase; 1.  
 KW Hydrolyase; Glycosidase.  
 SQ SEQUENCE 334 AA; 36440 MW; A50842168A944DAF CRC64;

Query Match 16.9%; Score 375; DB 3; Length 334;  
 Best Local Similarity 29.1%; Pred. No. 2e-18;  
 Matches 86; Conservative 55; Mismatches 125; Indels 30; Gaps 6;

QY 10 IPGIDFGCDIDGSCPTDTSVPLLSYKGGDAGQMKHFAEDDGLNFERISATQOFLVNT 69  
 DB 50 IPGVE---GDTYTPNTAIVL-----DQGNINIFRVPFLMRMMPNQ 90  
 QY 70 VDKLDELNMGSYNKVNACLETGATCMIDMNFARNGIIGCGVSDIFVDLWVQIA 129  
 DB 91 MTGVDVSAIYLOGYQVINYITSHGASAVIDPHNNGRYNNIIS---SPSDFQTFWNTIA 146  
 QY 130 KYEDNDKIIIFGLMNEPHDIEIMAQTCOKVYTAIRKAGATSOIILLPGTNFAVETVY 189  
 DB 147 SNFADNDGVIFDTNNEYHMDQDLVNLNOAALNGIRAGATSOIYFVEGNSWTGAMTW 206

DE Endoglucanase A.  
 GN EGLA.  
 OS Aspergillus kawachii.  
 OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;  
 OC Eurotiiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.  
 OK NCBI\_TaxID=40344;  
 RN (1)  
 RN SEQUENCE FROM N.A.  
 RA Ito K., Shimoi H., Iwashita K., Hara Y., Hinoki Y.;  
 RT "Cloning and characterization of endoglucanase genes from an  
 industrial fungus Aspergillus kawachii."  
 RL Submitted (FEB-2001) to the EMBL/Genbank/DBJ databases.  
 DR EMBL: AB053431; BAB62317.1;  
 DR InterPro: IPR001547; GH\_5.  
 DR InterPro: IPR001547; GH\_5.  
 DR Pfam: PF00150; cellulase.1.  
 DR Pfam: PF00150; cellulase.1.  
 DR PRODOM: PD001821; CBD\_fungal.1.  
 DR PROSITE: PS00562; CBD\_FUNGAL; UNKNOWN\_1.  
 SQ SEQUENCE 521 AA; 55034 MW; 34F23C719DA1A5EB CRC64;

Query Match 18.1%; Score 401.5; DB 3; Length 521;  
 Best Local Similarity 27.4%; Pred. No. 5.1e-20;  
 Matches 110; Conservative 70; Mismatches 170; Indels 51; Gaps 10;

QY 4 KYLVAIPGIDFGCDIDGSCPTD-----TSSVPLLSYKSGDAGQMKHFAEDDGLNFR 58  
 DB 32 KMWTSSESGAEFGSALPGTLGDTYTPETSKIQVLRNK-----GKNIFRI 76  
 QY 59 SATWQVYLVNTVDGKLDLNLMSYKVVNAACLETGAYCMIDMNFARYNGIIGGGVSD 118  
 DB 77 PLMERLTPDGLTSGFASVSTYSLDLKSTVEFVZNSGAVAVDPHNYGFRDGSII---EST 132  
 QY 119 DIPVDLWVOAKXYEDNKKIIFGLMNEPHDLEIMAGTQCKVYTAIRKAGTSQMLLP 178  
 DB 133 SDFKTMKVNATEFADNDKYIFDTNNEYHDMESLVNLNQAAINGRAAGATTQYTFVE 192  
 QY 179 GTNFASVETVYSTGSAEALGKITNPDGSTDLYFVHKYLDINNGSHACCTDNY--DA 236  
 DB 193 GNAVYTGAMDW--TTYNDLGLSLD---SEDKIIEVMHQVYLDSDSGSSECVSTICKER 247  
 QY 237 FNDFAWMLKONKQAIISFEGASMEPSCMTAFCAONKAISENSDYIIGVWGAGSEFDT 296  
 DB 248 IEKATWMLTKNNKQIIGFEGAGVNSVCEAVEGMLAYMSENSDVWGASWMSGPMWGT 307  
 QY 297 YILTEPLCKPGNYTDNKLMECIILOFTLDEYRPTPSISRAEETATATSDGDAR 356  
 DB 308 YWISLEP-----IDGTAYSTYL---PILKEYFPGSDSSSSASASVAAATG---AV 353  
 QY 357 STTKPIFREETASPTPNAVTKPSPDTSDDSDDDKDSASMS 397  
 DB 354 STTTAAFEQTTTPATQV-----EIASSSSSSSAAVAAST 388

RESULT 5  
 Q9PA12 PRELIMINARY; PRT; 356 AA.  
 AC Q9PA12;  
 DT 01-OCT-2000 (TREMBLrel. 15, Created)  
 DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)  
 DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)  
 DE Endo-1,4-beta-glucanase.  
 GN XF2708.  
 OS Xylella fastidiosa.  
 OC Bacteria; Proteobacteria; gamma subdivision; Xanthomonas group;  
 OC Xylella.  
 RN NCBI\_TaxID=2371;  
 RN (1)  
 RN SEQUENCE FROM N.A.  
 RC STRAIN=9ASC;  
 RX MEDLINE=20365717; PubMed=10910347;  
 RA Simpson A.J.G., Reimach F.C., Arruda P., Abreu F.A., Acencio M.,  
 Alvaranga R., Alves L.M.C., Araya J.E., Bala G.S., Baptista C.S.,

RA Barros M.H., Bonaccorsi E.D., Bordin S., Bove J.M., Briones M.R.S.,  
 RA Bueno M.R.P., Camargo A.A., Camargo L.E.A., Carraro D.M., Carier H.,  
 RA Colauto N.B., Colombo C., Costa F.F., Costa M.C.R., Costa-Neto C.M.,  
 RA Coutinho L.L., Cristofani M., Dias-Neto E., Docena C., El-Dorry H.,  
 RA Facincani A.P., Ferreira A.D.S., Ferreira V.C.A., Ferto J.A.,  
 RA Fraga J.S., Franca S.C., Franco M.C., Frohme M., Furlan L.R.,  
 RA Gartner M., Goldman G.H., Goldman M.H.S., Gomes S.L., Gruber A.,  
 RA Ho P.L., Hohnsels J.D., Junqueira M.L., Kemper E.L., Kitajima J.P.,  
 RA Krieger J.E., Kurmaev E.E., Laigret F., Lambis M.R., Leite L.C.C.,  
 RA Lemos E.G.M., Lemos M.V.F., Lopes S.A., Lopes C.R., Machado J.A.,  
 RA Machado M.A., Madelira A.M.B.N., Madelira H.M.F., Martino C.L.,  
 RA Marques M.V., Martins E.A.L., Martins E.M.F., Matsukuma A.Y.,  
 RA Menck C.F.M., Miracca E.C., Miyaki C.Y., Monteiro-Vitorello C.B.,  
 RA Moon D.H., Nagai M.A., Nascimento A.L.T.O., Netto L.E.S.,  
 RA Nandi A., Jr., Nobrega F.G., Nunes L.R., Oliveira M.A.,  
 RA de Oliveira M.C., de Oliveira R.C., Palmieri D.A., Paris A.,  
 RA Peixoto B.R., Pereira G.A.G., Pereira H.A., Jr., Pessegueiro J.B.,  
 RA Quaggio R.B., Roberto P.G., Rodrigues V., de Rosa A.J.M.,  
 RA de Rosa V.E., Jr., de Sa R.G., Santelli R.V., Sawasaki H.E.,  
 RA da Silva A.C.R., da Silva A.M., da Silva F.R., Silva W.A., Jr.,  
 RA de Souza A.P., Terenzi M.F., Truffi D., Tsai S.M., Tsuchino M.H.,  
 RA Valida H., Van Sluys M.A., Verjovski-Almeida S., Vettore A.L.,  
 RA Zago M.A., Zatz M., Zaidanis J., Zetubal J.C.,  
 RA "The genome sequence of the plant pathogen Xylella fastidiosa."  
 RL Nature 406:151-159(2000).  
 DR EMBL: AEO04077; AAF85505.1;  
 DR InterPro: IPR001547; GH\_5.  
 DR Pfam: PF00150; cellulase.1.  
 DR PROSITE: PS00659; GLYCOSYL\_HYDROL\_F5; UNKNOWN\_1.  
 SQ Complete proteome.  
 KW SEQUENCE 356 AA; 39273 MW; AA257413EC82C046 CRC64;

Query Match 17.5%; Score 387; DB 16; Length 356;  
 Best Local Similarity 29.3%; Pred. No. 3.1e-19;  
 Matches 92; Conservative 57; Mismatches 133; Indels 32; Gaps 8;

QY 3 IKYLVNIPGIDFGCDIDGSCPTDTSVPLLSK-----GGGAGQMKHFAEDDGLNFR 56  
 DB 37 LTYAGINLRPER-----SPSKRPGVLKQVLYPASPDAVYVK-----KGNIV 81  
 QY 57 RISATWQVYLVNTVDGKLDLNLMSYKVVNAACLETGAYCMIDMNFARYNGIIGGGV 116  
 DB 82 RLVLMEV--OPTLNDALDALTQCLKIIRKAVEQAKKLNIIIDVHNHYSEYNDDELIGDNY 140  
 QY 117 SDDIFVDLWQAIKXYEDNKKIIFGLMNEPHDLEIMAGTQCKVYTAIRKAGTSQML 176  
 DB 141 PISAFADLWMLKRLSLQFANDKAVIFGLMNEPHETSTTWAQAAQAAINAIKSTGACNNLV 200  
 QY 177 LPGTNFASVETVYST-----GSALGKITNPDGSTDLYFVHKYLDINNGSHAC--T 230  
 DB 201 VPGTAVSGAHSMWSSYYIGFNGEALNINP---AKHMAFEVQIYLNENSTGTGGEIST 257  
 QY 231 TDNVDAFNDFAWMLKONKQAIISFEGASMEPSCMTAFCAONKAISENSDYIIGVWGCA 290  
 DB 258 TIGAEEKLEAFTNMLRYHKFGFEGATGNDTCNALGEMLSYIEBNADVMGLTWGGS 317  
 QY 291 GS-FDTSYITLTP 303  
 DB 318 NPWFQPDYPPNLHP 331

RESULT 6  
 Q9C327 PRELIMINARY; PRT; 332 AA.  
 AC Q9C327;  
 DT 01-JUN-2001 (TREMBLrel. 17, Created)  
 DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)  
 DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)  
 DE Endoglucanase precursor.  
 GN ENGI.  
 OS Aspergillus niger.  
 OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;

```

QY 59 SATQFVLAIPIDGCDIDSCPTDTSVPLSYKGGDAGCGKHFPAEDDGLNFRISA 118
DB 153 PVGQFVLAIPIDGCDIDSCPTDTSVPLSYKGGDAGCGKHFPAEDDGLNFRISA 212
QY 119 DIFVDMVOIAKYEDNDKIIIFGLMNEPHDLDIEIWAOTCKQVYTAIRKAGATSMILLP 178
DB 213 AOFISLSQSLASKYASOSRWFEGLMNEPHDLDIEIWAOTCKQVYTAIRKAGATSMILLP 272
QY 179 TNEFASVEYVTSGAELALGKITNPDSSTDLIFDVHKYLDLNNSGSHAECTTDNV-AF 237
DB 273 GNDYQSAAFISDSGAALSGVTNPDSSTDLIFDVHKYLDLNNSGSHAECTTDNV-AF 332
QY 238 NDFADMLRQNRKRAIISSETGASMEPSCMTAFCAONKKAISESDVYIGFVGAGSFDTSY 297
DB 333 APLATYMLRQNRKRAIISSETGASMEPSCMTAFCAONKKAISESDVYIGFVGAGSFDTSY 392
QY 298 ILTLPGLKPGNYTDNKLMECI 320
DB 393 ILTLPGLKPGNYTDNKLMECI 415

```

## RESULT 2

```

ID 012665 PRELIMINARY; PRT; 410 AA.
AC 012665;
DT 01-NOV-1996 (TREMblrel. 01, Created)
DT 01-NOV-1996 (TREMblrel. 01, Last sequence update)
DE 01-JUN-2002 (TREMblrel. 21, Last annotation update)
DE Endoglucanase 2 (EC 3.2.1.4).
OS Egl2.
OC Penicillium janthinellum (Penicillium vitale).
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
OC Eurotiaceae; Trichocomaceae; mitosporic Trichocomaceae; Penicillium.
OX NCBI_TaxID=5079;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BIORGE IMET 43733.
RA MEDLINE=96207475; Pubmed=8625430.
RX Mennitz G., Koch A., Hentissat B., Schulz G.;
RT "Endoglucanase II (EglI) of Penicillium janthinellum - cDNA sequence,
RT heterologous expression and promoter analysis."
RL Curr. Genet. 29:490-495(1996).
DR EMBL: X89564; CA61740.1; -.
DR HSP: P00725; 2CBH.
DR InterPro: IPR000234; CBD_fungal.
DR InterPro: IPR001547; GH_5.
DR Pfam: PF00734; CBM_1; 1.
DR Pfam: PF00150; cellulase; 1.
DR ProDom: PD001821; CBD_fungal; 1.
DR SMART: SM00336; fcbd; 1.
DR PROSITE: PS00562; CBD_FUNGAL; 1.
DR PROSITE: PS00659; GLYCOSYL_HYDROL_F5; UNKNOWN_1.
KW Glycosidase; Hydrolase.
SQ SEQUENCE 410 AA; 43914 MW; 26B8395F3CD892A0 CRC64;

```

Query Match 40.0%; Score 885.5; DB 3; Length 410;  
 Best Local Similarity 50.0%; Pred. No. 5.6e-50;  
 Matches 163; Conservative 52; Mismatches 100; Indels 11; Gaps 3;

```

QY 1 GKIVLVAIPIDGCDIDSCPTDTSVPLSYKGGDAGCGKHFPAEDDGLNFRISA 60
DB 92 GKIVLVAIPIDGCDIDSCPTDTSVPLSYKGGDAGCGKHFPAEDDGLNFRISA 146
QY 61 TWQFVLNNTVVDGKLDLNMGSYNKRVNACLETGAYCMIDMNFARYNGIIGOGVSDDI 120
DB 147 GWOPIVNNNIGSLDSNNFKYEQVGSGLSGAYCIYDINHVARMGVIGOGVPDDO 206
QY 121 FVDLMVOIAKYEDNDKIIIFGLMNEPHDLDIEIWAOTCKQVYTAIRKAGATSMILLP 180
DB 207 FVSLMKQLATKYKNTKVAFCVNEPHDLDIEIWAOTCKQVYTAIRKAGATSMILLP 266
QY 181 NFASVEYVTSGAELALGKITNPDSSTDLIFDVHKYLDLNNSGSHAECTTDNV-AF 240
DB 393 ILTLPGLKPGNYTDNKLMECI 415

```

```

DB 267 DYSANFIENGSGAALPLVTNPDSSTDLIFDVHKYLDLNNSGSHAECTTDNV-AF 326
QY 241 ADMLRQNRKRAIISSETGASMEPSCMTAFCAONKKAISESDVYIGFVGAGSFDTS--YI 298
DB 327 ATWLSNRKRAIISSETGASMEPSCMTAFCAONKKAISESDVYIGFVGAGSFDTS--YI 386
QY 299 LTLPLKPGNYTDNKLMECI 324
DB 387 LTLPLKPGNYTDNKLMECI 408

```

## RESULT 3

```

ID 012637 PRELIMINARY; PRT; 368 AA.
AC 012637;
DT 01-NOV-1996 (TREMblrel. 01, Created)
DT 01-NOV-1996 (TREMblrel. 01, Last sequence update)
DT 01-MAR-2002 (TREMblrel. 20, Last annotation update)
DE Endo-beta-1,4-glucanase (EC 3.2.1.4).
OS Macrophoma phaseolina.
OC Eukaryota; Fungi; Basidiomycota; Macrophomina.
OX NCBI_TaxID=35725;
RN [1]
RP SEQUENCE FROM N.A.
RC MEDLINE=95309709; Pubmed=7789795;
RX Wang H., Jones R.W.;
RT "Cloning, characterization and functional expression of an
RT endoglucanase-encoding gene from the phytopathogenic fungus
RT Macrophoma phaseolina."
RL Gene 158:125-128(1995).
DR EMBL: U13914; AAB03889.1; -.
DR InterPro: IPR001547; GH_5.
DR Pfam: PF00150; cellulase; 1.
DR PROSITE: PS00659; GLYCOSYL_HYDROL_F5; UNKNOWN_1.
KW Glycosidase; Hydrolase.
SQ SEQUENCE 368 AA; 39945 MW; BA1FFB41119507CE CRC64;

```

Query Match 37.7%; Score 834; DB 3; Length 368;  
 Best Local Similarity 51.1%; Pred. No. 1.9e-50;  
 Matches 158; Conservative 48; Mismatches 99; Indels 4; Gaps 4;

```

QY 2 KIKVLAIPIDGCDIDSCPTDTSVPLSYKGGDAGCGKHFPAEDDGLNFRISA 61
DB 32 KIKVLAIPIDGCDIDSCPTDTSVPLSYKGGDAGCGKHFPAEDDGLNFRISA 90
QY 62 WQFVLNNTVVDGKLDLNMGSYNKRVNACLETGAYCMIDMNFARYNGIIGOGVSDDI 120
DB 91 WQFVLNNTVVDGKLDLNMGSYNKRVNACLETGAYCMIDMNFARYNGIIGOGVSDDI 150
QY 121 FVDLMVOIAKYEDNDKIIIFGLMNEPHDLDIEIWAOTCKQVYTAIRKAGATSMILLP 179
DB 151 FVSLMKQLATKYKNTKVAFCVNEPHDLDIEIWAOTCKQVYTAIRKAGATSMILLP 210
QY 180 TNEFASVEYVTSGAELALGKITNPDSSTDLIFDVHKYLDLNNSGSHAECTTDNV-DAFN 238
DB 211 NDWTSAAAFIDNGSAAALKKVTNPDSSTDLIFDVHKYLDLNNSGSHAECTTDNV-DAFN 270
QY 239 DPAFMLRQNRKRAIISSETGASMEPSCMTAFCAONKKAISESDVYIGFVGAGSFDTSY 298
DB 271 PLAWMLRQNRKRAIISSETGASMEPSCMTAFCAONKKAISESDVYIGFVGAGSFDTSY 330
QY 299 LTLPLKPG 307
DB 331 LVOTPISSP 339

```

## RESULT 4

```

ID 096WRO PRELIMINARY; PRT; 521 AA.
AC 096WRO;
DT 01-DEC-2001 (TREMblrel. 19, Created)
DT 01-DEC-2001 (TREMblrel. 19, Last sequence update)
DE 01-JUN-2002 (TREMblrel. 21, Last annotation update)

```

GenCore version 5.1.4-p5.4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 9, 2003, 15:15:41 ; Search time 102.358 Seconds  
(without alignments)  
843.445 Million cell updates/sec

Title: US-10-028-245-2

Perfect score: 2214

Sequence: 1 GKIKYLGVAIPGIDFGCDID.....GLTGTVLFTVAALGYMLVAF 419

Scoring table:

BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

SPTREMBL\_21:\*  
1: sp\_archaea:\*  
2: sp\_bacteria:\*  
3: sp\_fungi:\*  
4: sp\_human:\*  
5: sp\_invertebrate:\*  
6: sp\_mammal:\*  
7: sp\_mmc:\*  
8: sp\_organelle:\*  
9: sp\_phage:\*  
10: sp\_plant:\*  
11: sp\_protent:\*  
12: sp\_virus:\*  
13: sp\_vertebrate:\*  
14: sp\_unclassified:\*  
15: sp\_virus:\*  
16: sp\_bacteriophage:\*  
17: sp\_archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Match	Query Length	DB ID	Description
1	944.5	42.7	418	093833	093833 trichoderma
2	885.5	40.0	410	012665	012665 penicillium
3	834	37.7	368	012637	012637 macrophomina
4	401.5	18.1	521	096WRO	096WRO aspergillus
5	387	17.5	356	16	09PA12
6	381.5	17.2	332	09C327	09C327 xylophora
7	379.5	17.1	332	096WRO	096WRO aspergillus
8	375	16.9	334	08WZD7	08WZD7 talaromyces
9	371.5	16.8	389	09C328	09C328 volvariella
10	366.5	16.6	412	074169	074169 aspergillus
11	361	16.3	331	074169	074169 aspergillus
12	354	16.0	332	059951	059951 aspergillus
13	349.5	15.8	335	08TG26	08TG26 thermascus
14	346	15.6	385	09P981	09P981 robillarda
15	345.5	15.6	325	09Y8H6	09Y8H6 emericella
16	345.5	15.6	364	09F0G8	09F0G8 rhizobium m

17	339.5	15.3	335	3	096U15	096U15 thermascus
18	337.5	15.2	333	3	012638	012638 macrophomina
19	221	10.0	229	2	093GB3	093GB3 raistoria s
20	219	9.9	229	2	093GB5	093GB5 raistoria s
21	219	9.9	229	2	093GB4	093GB4 raistoria s
22	219	9.9	229	2	093GB2	093GB2 raistoria s
23	219	9.9	229	2	093GB9	093GB9 raistoria s
24	219	9.9	229	2	093GB8	093GB8 raistoria s
25	219	9.9	231	2	093GB4	093GB4 raistoria s
26	219	9.9	231	2	093GB4	093GB4 raistoria s
27	217	9.8	231	2	093GA7	093GA7 raistoria s
28	215	9.7	231	2	093GB1	093GB1 raistoria s
29	214	9.7	231	2	093GA4	093GA4 raistoria s
30	212	9.6	231	2	0932X5	0932X5 raistoria s
31	210	9.5	231	2	093GA9	093GA9 raistoria s
32	210	9.5	231	2	093GA3	093GA3 raistoria s
33	210	9.5	231	2	093GA2	093GA2 raistoria s
34	210	9.5	231	2	093AN2	093AN2 raistoria s
35	210	9.5	231	2	0933E6	0933E6 raistoria s
36	209	9.4	231	2	093GB0	093GB0 raistoria s
37	208	9.4	231	2	093GA6	093GA6 raistoria s
38	203	9.2	231	2	093GA5	093GA5 raistoria s
39	201	9.1	231	2	093GA1	093GA1 raistoria s
40	186	8.4	1232	3	059943	059943 neocallimias
41	175	7.9	471	3	P78719	P78719 orpionomyces
42	170.5	7.7	477	3	013333	013333 orpionomyces
43	163.5	7.4	477	3	09HFE0	09HFE0 orpionomyces
44	163.5	7.4	482	3	001409	001409 neocallimias
45	159	7.2	362	3	09URH5	09URH5 piromyces r

## ALIGNMENTS

## RESULT 1

093833 PRELIMINARY: PRT: 418 AA.  
ID 093833;  
AC 01-MAY-1999 (TREMBLrel. 10, Created)  
DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)  
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)  
DE Endoglucanase II.  
EGL2.  
OS Trichoderma viride.  
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;  
OC Hypocreales; Mitosporic Hypocreales; Trichoderma.  
OX NCBI\_TaxID=5547;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=MC300-1;  
RA Watanabe M.;  
RT "Endoglucanase II-T. viride."  
RL Submitted (DEC-1998) to the EMBL/Genbank/DBJ databases.  
DR EMBL; AB021657; BAA36216.1; -  
DR HSP; P00725; 2CBH.  
DR InterPro: IPR000254; CBD\_fungal.  
DR InterPro: IPR001547; GH\_5.  
DR Pfam: PF00734; CBM\_1; 1.  
DR Pfam: PF00150; cellulase; 1.  
DR ProDom: PD001821; CBD\_fungal; 1.  
DR SMART; SM00236; fcbd; 1.  
DR PROSITE; PS00562; CBD\_FUNGAL; 1.  
DR PROSITE; PS00659; GLYCOSYL\_HYDROL\_F5; UNKNOWN; 1.  
SQ SEQUENCE 418 AA; 44154 RM; BC193E24B57F250B CRC64;

Query Match 42.7%; Score 944.5; DB 3; Length 418;  
Best local similarity 52.3%; Pred. No. 4.4e-58;  
Matches 169; Conservative 57; Mismatches 92; Indels 5; Gaps 2;

QY 3 IKYLGVAIPGIDFGCDIDSCPTDITSSVPLSTKGG---DAGQKHRAEDGLVAFRI 58  
DB 93 VRAGVNIAGFDGCTDGTCTSKVYPLPKETGANNYPDGIQGMQHVNDDGKMTIFRL 152

**THIS PAGE BLANK (USPTO)**

OY 317 NECILOFTLDEKXRPPTSTAEETATATASDGDAPSITKPIFREETASPTNAV 376  
Db 342 VAGKAETFGIFNR-----SNLTWDAPVMAKAFIKGIGS-SITTP--TTPTPTPTPTPT 393  
OY 377 KPSPDTSDDSDDDKDSASMS 397  
Db 394 TPTPTPTPTPTPTPOSANEVT 414

Search completed: May 9, 2003, 15:18:53  
Job time : 18.2626 secs



```

ID GUNA_RUMAL STANDARD: PRT: 364 AA.
AC P23660:
DT 01-NOV-1991 (Rel. 20, Created)
DT 01-NOV-1991 (Rel. 20, Last sequence update)
DT 01-NOV-1994 (Rel. 29, Last annotation update)
DE Endoglucanase A (EC 3.2.1.4) (Endo-1,4-beta-glucanase) (cellulase)
DE (ECA).
GN CELA.
OS Ruminococcus albus.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Lachnospiraceae;
OC Ruminococcus.
OX NCBI_TaxID=1264;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 1-14.
RC STRAIN-SY3.
RX MEDLINE-91066833; PubMed-2250649;
RA Poole D.M., Hazlewood G.P., Laurie J.I., Barker P.J., Gilbert H.J.;
RT "Nucleotide sequence of the Ruminococcus albus SY3 endoglucanase
RT genes cels and celsB."
RL Mol. Gen. Genet. 223:217-223(1990).
CC -1- FUNCTION: HYDROLYSES BOTH CARBOXYMETHYLCELLULOSE AND XYLAN.
CC PROBABLY HAS A ROLE IN HYDROLYZING OLIGOSACCHARIDES DERIVED
CC FROM CELLULOSE, WHICH ARE TRANSPORTED ACROSS THE CELL WALL.
CC -1- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic
CC linkages in cellulose.
CC -1- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-xylosidic
CC linkages in xyans.
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (Probable).
CC -1- SIMILARITY: BELONGS TO CELLULOSE FAMILY A (FAMILY 5 OF GLYCOSYL
CC HYDROLASES).
CC -----
CC DR EMBL: X54931; CAA38692.1: -
CC DR PIR: S12017; S12017.
CC DR HSSP: P17901; 1EDG.
CC DR InterPro: IPR001547; GH_5.
CC DR Pfam: PF00150; cellulase; 1.
CC DR PROSITE: PS00659; GLYCOSYL_HYDROL_F5; 1.
CC DR Cellulose degradation: Hydrolase; Glycosidase; Xylan degradation.
CC ACT_SITE 169 169 PROTON DONOR (BY SIMILARITY).
CC FT ACT_SITE 293 293 NUCLEOPHILE (BY SIMILARITY).
CC FT ACT_SITE 364 AA; 41218 MW; F8DCDDE2E55781F7 CRC64;
CC SQ SEQUENCE
CC -----
Query Match 6.1%; Score 136; DB 1; Length 364;
Best Local Similarity 21.0%; Pred. No. 0.0099;
Matches 66; Conservative 54; Mismatches 119; Indels 76; Gaps 14;

```

```

OY 266 TAFCAONKAISEND 280
DB 292 TEYGAIVK---DNND 303
RESULT 14
GUNA_CLOCE
ID GUNA_CLOCE STANDARD: PRT: 475 AA.
AC P17901;
DT 01-NOV-1990 (Rel. 16, Created)
DT 01-NOV-1990 (Rel. 16, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE Endoglucanase A precursor (EC 3.2.1.4) (Endo-1,4-beta-glucanase A)
DE (Cellulase A) (EGCCA).
GN CELCA.
OS Clostridium cellulolyticum.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1521;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 27-39.
RC STRAIN-ATCC 35319;
RX MEDLINE-90108715; PubMed-2558058;
RA Faure E., Belatch A., Bagnara C., Gaudin C., Belatch J.-P.;
RT "Sequence analysis of the Clostridium cellulolyticum endoglucanase-A-
RT encoding gene, celCCA."
RL Gene 84:39-46(1989).
RN [2]
RN CHARACTERIZATION.
RX MEDLINE-92078105; PubMed-1744052;
RA Fierobe H.-P., Gaudin C., Belatch A., Loufti M., Faure E., Bagnara C.,
RA Baty D., Belatch J.-P.;
RT "Characterization of endoglucanase A from Clostridium
RT cellulolyticum."
RL J. Bacteriol. 173:7956-7962(1991).
RN [3]
RN MUTAGENESIS OF ARG-104; HIS-147 AND HIS-148.
RX MEDLINE-92325059; PubMed-1624455;
RA Belatch A., Fierobe H.-P., Baty D., Busetta B., Bagnara-Tardif C.,
RA Gaudin C., Belatch J.-P.;
RT "The catalytic domain of endoglucanase A from Clostridium
RT cellulolyticum: effects of arginine 79 and histidine 122 mutations on
RT catalysis."
RL J. Bacteriol. 174:4677-4682(1992).
RN [4]
RN X-RAY CRYSTALLOGRAPHY (1.6 ANGSTROMS) OF 27-406.
RP STRAIN-ATCC 35319;
RX MEDLINE-96097400; PubMed-8535787;
RA Ducros V., Czjzek M., Belatch A., Gaudin C., Fierobe H.-P.,
RA Belatch J.-P., Davies G.J., Haer R.;
RT "Crystal structure of the catalytic domain of a bacterial cellulase
RT belonging to family 5."
RL Structure 3:939-949(1995).
CC -1- FUNCTION: THE BIOLOGICAL CONVERSION OF CELLULOSE TO GLUCOSE
CC GENERALLY REQUIRES THREE TYPES OF HYDROLYTIC ENZYMES:
CC (1) EXOCELLULOBRANOLYSES WHICH CUT INTERNAL BETA-1,4-GLUCOSIDIC BONDS;
CC (2) EXOCELLULOBRANOLYSES THAT CUT THE DISSACCHARIDE CELLULOSE
CC FROM THE NONREDUCING END OF THE CELLULOSE POLYMER CHAIN;
CC (3) BETA-1,4-GLUCOSIDASES WHICH HYDROLYZE THE CELLULOSE AND OTHER
CC SHORT CELLO-OLIGOSACCHARIDES TO GLUCOSE.
CC -1- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic
CC linkages in cellulose.
CC -1- DOMAIN: A 24 RESIDUES DOMAIN IS REPEATED TWICE IN THIS ENZYME AS
CC WELL AS IN OTHER C.THERMOCELLUM CELLULOSE ENZYMES. THIS DOMAIN
CC MAY FUNCTION AS THE BINDING LIGAND FOR THE SL COMPONENT.
CC -1- MISCELLANEOUS: THE C-TERMINUS (AA 411-475) MAY PLAY A ROLE IN
CC ORGANIZING THE CELLULOSE COMPLEX.
CC -1- SIMILARITY: BELONGS TO CELLULOSE FAMILY A (FAMILY 5 OF GLYCOSYL
CC HYDROLASES).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

```

```

DR InterPro: IPR001547; GH_5.
DR Pfam: PF00150; cellulase; 1.
DR Pfam: PF00553; CBM_2; 1.
DR PROSITE: PS00659; GLYCOSYL_HYDROL_F5; 1.
KW Cellulose degradation: Hydrolyase; Glycosidase; Signal.
FT SIGNAL 1 25
FT CHAIN 26 517
FT DOMAIN 26 517
FT ACT_SITE 421 517
FT ACT_SITE 185 185
FT ACT_SITE 309 309
SQ SEQUENCE 517 AA; 57660 MM; AID1570302PFB30 CRC64;

Query Match
Best Local Similarity 21.2%; Score 141; DB 1; Length 517;
Matches 94; Conservative 70; Mismatches 173; Indels 106; Gaps 23;

OY 3 IKYLVAVIPGIDPCDDIDGSCPTPTSSVPLSYKGGDC---ACQMKHFAEDDGLNFR 58
DB 43 VOEWGV---GWNLDGNTLDKATNLSTYNSPISFETGWCNPTVTAKMDIKINAGFKTIRI 99
OY 59 SATWQFVLNNTVDG--KLDELINMGSYKNVYACLETGAYCMIDHNFARYNGIIGQGV 116
DB 100 PTTW---GHLGDNKNNKNEEMVARKVEVVDYCIADLYVLNTHHEGNVYIPYAKSS 155
OY 117 SDDIFVLWVOIAKYED-NDKIIFGLMNEP--HDLDIEIWAQTCQK-----VVT 163
DB 156 VTPRLKTLMTQISEAFKDYDHLIFETLNEPRELCTPEWMTGTSSESHVYKNYNAALE 215
OY 164 AIRKAGAT--SOMILLPGTNASVETVSTGSAALCKITPDOSTDLYLDVYKYL--- 218
DB 216 SIKRTGNNLSRAVAMP-----TYAAGSSSTTMDKVPDKKN--VIASVHAYSPYE 265
OY 219 -----DINSGSHAECTTDYDAFNDAFMDLRONKRAIISSETG-----ASM 260
DB 266 FAMDTSSNVWTGSSVYDKSLDVE-LDSYLNTEFKSKGVPIVIEFGSINKNNTSSRAEL 324
OY 261 EPSQMTAFCAQNKRAISENSDYVIGFVNGAGSEPTSTILVLPJCKPGNYTDNK-----L 315
DB 325 AEYVYTA--AKRCRIP-----CVMW-----DNNYAEI-----NKGEFGL 357
OY 316 MNECILQFTLDEKYRPTPTISIAEETAATATSDGADAST--TKPIREETASPTPN 373
DB 358 LNKSTLWMTYSDIK-----DALIRGYKNVHPEATED-DKSTVUTNP---DSGNTRPD 406
OY 374 A-VTKPSPTSDSDDDDKDASAS 395
DB 407 SGNTPCTETTTPTDNEKISITS 429

RESULT 12
GUN5_BACAG STANDARD; PRT; 400 AA.
AC 085465;
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE Endoglucanase 5A (EC 3.2.1.4) (Endo-1,4-beta-glucanase) (Alkaline
DE cellulase).
GN CEL5A.
OS Bacillus agaradhaerens (Bacillus agaradhaerans).
OC Bacteria; Firmicutes; Bacilliales; Bacillaceae; Bacillus.
OX NCBI_TaxID=76935;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=DSM 8721;
RA Bjornvad M.E.;
RL Submitted (MAY-1998) to the EMBL/GenBank/DBJ databases.
RN [2]
RP X-RAY CRYSTALLOGRAPHY (1.57 ANGSTROMS) OF 30-329.
RC STRAIN=AC13 / NCIMB 40482;
RX MEDLINE=98153671; PubMed=9485319;
RA Davies G.J., Dauter M., Brzozowski A.M., Bjornvad M.E.,

```

```

RA Andersen K.V., Schuelein M.;
RT "Structure of the Bacillus agaradhaerans family 5 endoglucanase at 1.6-
RT A and its cellobiose complex at 2.0-A resolution.";
RL Biochemistry 37:1926-1932(1998).
RN [3]
RP X-RAY CRYSTALLOGRAPHY (1.64 ANGSTROMS) OF 30-329.
RC STRAIN=AC13 / NCIMB 40482;
RX MEDLINE=98384136; PubMed=9718293;
RA Davies G.J., Mackenzie L.F., Varrot A., Dauter M., Brzozowski A.M.,
RA Schuelein M., Withers S.G.;
RT "Snapshots along an enzymatic reaction coordinate: analysis of a
RT retaining beta-glycoside hydrolase.";
RL Biochemistry 37:11707-11713(1998).
CC -1- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic
CC linkages in cellulose.
CC -1- SUBUNIT: MONOMER.
CC -1- SIMILARITY: BELONGS TO CELLULASE FAMILY A (FAMILY 5 OF GLYCOSYL
CC HYDROLASES).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AF067428; AAC19169.1; -
DR PDB: 1A3H; 16-MAR-99.
DR PDB: 2A3H; 16-MAR-99.
DR PDB: 3A3H; 16-MAR-99.
DR InterPro: IPR003610; CBM_5_12.
DR InterPro: IPR001547; GH_5.
DR Pfam: PF00150; cellulase; 1.
DR Pfam: PF002839; CBM_5_12; 1.
DR SMART: SM00495; ChBD3; 1.
DR PROSITE: PS00659; GLYCOSYL_HYDROL_F5; 1.
KW Cellulose degradation: Hydrolyase; Glycosidase; 3D-structure.
FT ACT_SITE 165 165
FT ACT_SITE 254 254
SQ SEQUENCE 400 AA; 44702 MM; 3F9C66FB9BC36FFF CRC64;

Query Match
Best Local Similarity 19.9%; Score 136.5; DB 1; Length 400;
Matches 70; Conservative 60; Mismatches 141; Indels 81; Gaps 14;

OY 44 MKHFAEDDGLVFR---ISATWQFVLNNTVDGKIDELINMGSYKNVYACLETGAYCMIDM 100
DB 75 MKWLRLDGMIVFRAAMYTSSGYIDDPVSKYKE-----AVEAIDLDIYIYIDW 126
OY 101 HNFARYNGIIGQGVSDIDIVDLWVOIAKYEDNDKIIFGLMNEPHLDIEIWAQTCQK--- 157
DB 127 HILSDNPNITKEEA-----KDFEDENSELYGDPVNYIEIANEPNSDV--TNGNQIKP 179
OY 158 -CQRYVTAIRKAGATSQMILLPGTNFASVETVSTGSAALCKITPDOSTDLYLDVYK 216
DB 180 YAEVYIPIRNDPNN--IIVGTGWSQDVHNA-----ADNQLADPN-----VMAVHFE 227
OY 217 YLDLNNGSHAECTTDYDAFNDAFMDLRONKRAIISSETGASMEPSQMTAFCAQNKRAIS 276
DB 228 Y-----AGTHGQNLRDY-----DYALDQGAALFVSEMGTS-----AATGDGCVFL 268
OY 277 ENSDVYIGEV-----GWCAGSFDSYILTLPLCKPGVYTNKLMNECILQFTLDE 328
DB 269 DEAYWIDFMDERLNLNANMSLTIRKDESSAALMGANGNTGWTAEEL----- 315
OY 329 KYRPTPTISIAEETAT-----ATATSDGAPSTPTPIREETASPTPNVAV 375
DB 316 --SPSGFVREKIRASASIPSPDPTPSDPEPDPPTPSDGEVPAMDPNQI 365

RESULT 13
GUNA_RUMAL

```

```

Db 150 YVPPNN--ANKORSEKFFYSIMTQIAKEFKKNYHLYHFETMNEPRVLGHG---EEWFFPR 203
Oy 155 -----AQC-----OKVYAIRKAGA--TSOMILLPGTNEFASVETVSTGSAEALG 198
Db 204 NNPNSNDIREVACINDINYNOVALAIRATGNNATRCVMPFGYD-ASIEGCMTGCF----- 257
Oy 199 KITNPDGSDLLYEDVHKYLDIN--NSGSHAECTDN--VDA-FNDFADWLROKRRQ 250
Db 258 KMPN-DTASGRLLSYHAYIPYEFALASDTYVFRPDNLKDYIDSEFNDLSNFKLRNP 316
Oy 251 AIISFEGASMEPSCMTAFCAQONKAISEN-SDVYIGFVGAGSGFDTSYITLTPLGKPGN 309
Db 317 VVVGETSATNR-----NNTAERYKNADY-----NGRARSYNAVYL----- 354
Oy 310 YTDNKLAM-----NEC--ILDQFTLDEKRYRPTSTISAAEETATATATSDG-DAPST 358
Db 355 -WDNNTYONNSAGSDECHMYIDRNSLQWK--DPEITISTMKHVDTGPTATNGKEIPST 410
Oy 359 TKPIREETASPTNPAVTRKPSPTSDSDDDKD 391
Db 411 EGDPTPVDDPTPR-VDPTPTPVDDPTPVDDPD 442

```

## RESULT 10

```

GUN2_BACS4
ID GUN2_BACS4 STANDARD; PRT; 409 AA.
AC P06565;

```

```

Dt 01-JAN-1988 (Rel. 06, Created)
Dt 01-JAN-1988 (Rel. 06, Last sequence update)
Dt 16-OCT-2001 (Rel. 40, Last annotation update)
DE Endoglucanase B (EC 3.2.1.4) (Endo-1,4-beta-glucanase) (Cellulase).
GN CELB.
OS Bacillus sp. (strain N-4 / JCM 9156).
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=1413;
RN [1]
RP MEDLINE=87056924; PubMed=3782013;
RA Fukumori F., Sashihara N., Kudo T., Horikoshi K.;
RT "Nucleotide sequences of two cellulase genes from alkalophilic
RL Bacillus sp. strain N-4 and their strong homology."
J. Bacteriol. 168:479-485(1986).
CC -1- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic
links in cellulose.
CC -1- SIMILARITY: BELONGS TO CELLULASE FAMILY A (FAMILY 5 OF GLYCOSYL
HYDROLASES).

```

```

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).

```

```

CC EMBL, M14729; AAA22299.1; -
DR PIR: B25156; B25156.
DR HSSP: O85465; 1A3H.
DR InterPro: IPR003610; CBM_5_12.
DR InterPro: IPR001547; GH_5.
DR Pfam: PF002839; CBM_5_12; 1.
DR SMART: SM00495; CHBD3; 1.
DR PROSITE: PS00659; GLYCOSYL_HYDROL_F5; 1.
KM Cellulose degradation; Hydrolase; Glycosidase.
FT ACT_SITE 165
FT ACT_SITE 254
FT ACT_SITE 254 PROTON DONOR (BY SIMILARITY).
FT ACT_SITE 254 NUCLEOPHILE (BY SIMILARITY).
SO SEQUENCE 409 AA; 45690 MW; 9888660E6B4D3F CRC64;

```

```

Query Match 6.6%; Score 147; DB 1; Length 409;
Best Local Similarity 20.9%; Pred. No. 0.0019;
Matches 75; Conservative 59; Mismatches 139; Indels 86; Gaps 16;

```

```

Oy 44 MKHFAEDDGLNVER---ISATMOFVLNNTVDGKLDELNMGSYNKRYVNAACLETCAYCMTM 100
Db 75 MKMLRDMQWGITVRAAMYTSSGGYIEDPSYKEVKE-----AYEAIDLGIYITDM 126
Oy 101 HNFARINGGITGGGVSDIFVDLWQIAKYEDNDKIIIFGLMNEPHDLDEITMAOT--- 157
Db 127 HILSDNDPNLYKEEA-----KDFEEMSELYGDYPAVYIEALNEPNSGV-TWMDQIKP 179
Oy 158 -CQKVYAIKPKATGSMILLPGTNEFASVETVSTGSAEALGITNPDGSDLLYEDVHK 216
Db 180 YAEELVPIKNDPN--IITVGTGWSQDVHNA-----ADNLTDPN-----VMAFHF 227
Oy 217 YLDINNSGSHAECTTNDVDAFNDFADWLROKROAISETGASMEPSCMTAFCAQONKAIS 276
Db 228 Y-----AGTHGQULNRQV-----DYALDQGAIFPSENGTSEATNGDGVFL----- 268
Oy 277 ENSDVIIGEV-----GKAGSFDTSYITLTPLGKPGNYVDNKLKMECIIIDQFTLDE 328
Db 269 DEAQWVIDFMDERNLGSMANSLTHKDESSAALMPGASPTGWTAEVL----- 315
Oy 329 KYRPTSTISAAEETATATATSDGAPSTTKPIFREETASPTNPAVTRKPSPTSDSD 387
Db 316 -SPSGTEVREKIRESAT-TPSPDPTPVDDPD-----GEPPD--PEPPTPVDDPD 362

```

## RESULT 11

```

GUN2_CLOLO
ID GUN2_CLOLO STANDARD; PRT; 517 AA.
AC P54937;

```

```

Dt 01-OCT-1996 (Rel. 34, Created)
Dt 01-OCT-1996 (Rel. 34, Last sequence update)
Dt 01-OCT-1996 (Rel. 34, Last annotation update)
DE Endoglucanase A precursor (EC 3.2.1.4) (Endo-1,4-beta-glucanase A)
GN CELA.
OS Clostridium longisporum.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OX NCBI_TaxID=1523;
RN [1]
RP MEDLINE=94172316; PubMed=8126442;
RA Mittenfior V., Thomson J.A.;
RT "Cloning of an endo-(1->4)-beta-glucanase gene, celsA, from the rumen
RT bacterium Clostridium sp. ('C. longisporum') and characterization of
RT its product, CelA, in Escherichia coli."
J. Gen. Microbiol. 139:3233-3242(1993).
CC -1- FUNCTION: THE PH AND TEMPERATURE OPTIMA OF CELA ARE 4.8 AND 4.3
CC DEGREES CELSIUS, RESPECTIVELY. IT HYDROLYSES BARLEY BETA-GLUCAN,
CC LICHENAN, CARBOXYMETHYLCELLULOSE AND XYLAN. IT SHOWS PREFERENCE FOR
CC ACTIVITY AGAINST THE LARGER CELLULOSE POLYSACCHARIDES (CELLOHEXAPOSE
CC AND CELLOPENTAPOSE); CELLOTETRAPOSE IS THE SMALLEST SUBSTRATE
CC DEGRADED COMPLETELY.
CC -1- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic
CC linkages in cellulose.
CC -1- SIMILARITY: BELONGS TO CELLULASE FAMILY A (FAMILY 5 OF GLYCOSYL
CC HYDROLASES).
CC -1- SIMILARITY: CONTAINS 1 BACTERIAL-TYPE CELLULOSE-BINDING DOMAIN
(CBD).

```

```

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).

```

```

CC EMBL: L02868; AAC37035.1; -
DR HSSP: P17901; IEDG.
DR InterPro: IPR001919; Bac_cellose-bind.

```



```

CC -1- PATHWAY: Cellulose degradation.
CC -1- SIMILARITY: BELONGS TO CELLULOSE FAMILY A (FAMILY 5 OF GLYCOSYL
CC HYDROLASES).
CC -1- CAUTION: THIS SEQUENCE IS PROBABLY INCORRECT AND FRAMESHIFTING
CC IT IN REGION 124-163 WILL INCREASE THE SIMILARITY WITH THE OTHER
CC MEMBERS OF THIS FAMILY (AND RESTORE AN ACTIVE SITE RESIDUE!).
CC
CC PIR: JX0131.
CC InterPro: IPR001547; GH_5.
CC Pfam: PF00150; cellulase: 1.
CC DR PROSITE: PS00659; GLYCOSYL_HYDROL_F5; FALSE_NEG.
CC Cellulose degradation: Hydrolase; Glycosidase; Glycoprotein.
CC FT CAROHD 76 76 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC FT CAROHD 182 182 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC FT CAROHD 271 271 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC FT ACT_SITE 266 266 NUCLEOPHILE (BY SIMILARITY).
CC SO SEQUENCE 375 AA; 40870 MW; ALDD7C186AABEE9 CRC64;

Query Match 9.8%; Score 217.5; DB 1; Length 375;
Best Local Similarity 24.9%; Pred. No. 1.3e-08;
Matches 82; Conservative 50; Mismatches 128; Indels 69; Gaps 14;

OY 3 IKVLGVAIFGIDPGCDIDSCPTDTSSVPLLSYKGGDA---GOKKHFADDDGLAVERT 58
DB 2 IYAGVAAENGNGERV---WSATOTPGTGLPGRFQVDYAFISAADVAVDQNHLLERFV 57
OY 59 SATWQFVLNNT-----VDGKLDELWMSYKYNVNACTET-GAYCMIDMNFARVN--- 107
DB 58 A---FLERMCPPATGLAAAFNEHFDEYKFAVDYITVKGAVAILDPHNYRVNVLAT 113
OY 108 -----GGIIGGGVSDDIFFDLW-----VOIAYIYENDKIIITGLMN 144
DB 114 SLSEVSLATPOTPOPPOPSLSAFGV-----WHLASMTSAHFHERATRG--- 163
OY 145 EPHDDIEIACQCKVYTAIRKAGATSOIILPGTNEFASVETVV---STGSAFALKITN 202
DB 164 -----DLVVLANN-QAATAIRAANA-SNLIIPGNSWTGSHGTSDESSALLWQFKD 216
OY 203 PGGSTDLLEYVHKYLDINNNGSHACTDNDVAFNDVAFWLEQNKROAIISGTASMEP 262
DB 217 PLNNTAI---DIHEVLDYDFSGHLECVSPETINLAALMLKENLKAFTTEGGSNST 273
OY 263 SCRTAFCAONKAISENSDYITGVGAG 291
DB 274 SCQEMLPDLINYMADNAE-YIGWTAWAG 301

RESULT 7
GUNE_CLOTM
ID GUNE_CLOTM STANDARD: PRT; 814 AA.
AC P10477;
DT 01-JUL-1989 (rel. 11, Created)
DT 01-JUL-1989 (rel. 11, Last sequence update)
DT 01-FEB-1995 (rel. 31, Last annotation update)
DE Endoglucanase E precursor (EC 3.2.1.4) (EGE) (Endo-1,4-beta-glucanase)
GN CELE.
OS Clostridium thermocellum.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxId=1515;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 35-58.
RX MEDLINE=89137992; PubMed=3066698;
RA Hall J., Hazlewood G.P., Barker P.J., Gilbert H.J.;
RT "Conserved reiterated domains in Clostridium thermocellum
RT endoglucanases are not essential for catalytic activity.";
RL Gene 69:29-38(1988).
CC -1- FUNCTION: THIS ENZYME CATALYZES THE ENDOHYDROLYSIS OF 1,4-BETA-
CC GLUCOSIDIC LINKAGES IN CELLULOSE, LICHENIN AND CEREAL BETA-D-
CC GLUCANS.
CC -1- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic
CC linkages in cellulose.
CC -1- DOMAIN: A 24 RESIDUES DOMAIN IS REPEATED TWICE IN THIS ENZYME AS

```

```

CC WELL AS IN OTHER C.THERMOCCELLUM CELLULOSOME ENZYMES. THIS DOMAIN
CC MAY FUNCTION AS THE BINDING LIGAND FOR THE SL COMPONENT.
CC -1- SIMILARITY: BELONGS TO CELLULOSE FAMILY A (FAMILY 5 OF GLYCOSYL
CC HYDROLASES).
CC
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: M22759; AAA23224.1; -
CC PIR: JTO347; CCELEM.
CC HSSP: P17901; 1EDG.
CC
CC DR InterPro: IPR002105; Dockerin_1.
CC DR InterPro: IPR001547; GH_5.
CC DR InterPro: IPR001087; Lipase_GDSL.
CC Pfam: PF00150; cellulase: 1.
CC DR Pfam: PF00404; Dockerin_1; 2.
CC DR PROSITE: PS00448; CLOS_CELLULOSOME_RPT; 2.
CC DR PROSITE: PS00659; GLYCOSYL_HYDROL_F5; 1.
CC Cellulose degradation: Hydrolase; Glycosidase; Repeat; Signal.
CC KW SIGNAL 1 34
CC FT CHAIN 35 814
CC FT ACT_SITE 193 193 PROTON DONOR (BY SIMILARITY).
CC FT ACT_SITE 316 316 NUCLEOPHILE (BY SIMILARITY).
CC FT DOMAIN 415 474 2 X 24 AA APPROXIMATE REPEATS.
CC FT REPEAT 415 438
CC FT REPEAT 451 474
CC SO SEQUENCE 814 AA; 90244 MW; C6FA24BBD1523632 CRC64;

Query Match 8.7%; Score 193.5; DB 1; Length 814;
Best Local Similarity 23.5%; Pred. No. 2.1e-06;
Matches 92; Conservative 60; Mismatches 157; Indels 83; Gaps 19;

OY 52 GLVFRISATWQFVLNNTVVGKLDLWMSYKYNVNACTETGAYCMIDMNFARVNGIT 111
DB 99 GFNAVRPPTWDRHIGPADPYKIDEAMLNREVEVNVYLDGKATIIINHDTWIIPT 158
OY 112 GGGVSDDIFFDLWQIAKYED-NDKIIIFGLNEPHDL--DIEIWAQTCQK----- 160
DB 159 ANEORSKEREKLVKYMEDIAATFRKDYDHLLEFETNNEPREVSGPHEMGGTENRDVYNRFN 218
OY 161 --VVTAIRKAGATSOIILPGTNEFASVETVVSGSABALKITNPGGSTDLLEYVHKY- 217
DB 219 LAVVNTIRASGNN-----KRFILVPTNAATGLVALNDLVTPNNDSRFV-VSIHAYS 271
OY 218 -----LDINNS--GSHAE--CTTDNVDA-FNFDWMLRQNKROAIISGTASMEPSCWT 266
DB 272 PYFFAMDVNGTSTWGSDDYDASLTSELDAIYNFV---KNGAVIIGERG-TIDKNNIS 326
OY 267 AFCAQ-----NKAISENSDYITGVGAGSFTSYILTLPLKPGNTDNKLMECIL 321
DB 327 SRVAHAHVARAVSGLAVF-----W-----WNGVY-----NNGDAFTVALLRKTL 370
OY 322 DDF-----TLDEKYRPTT-SISTAAEEFATAT-----ATSGDAPSTTKPIF 363
DB 371 SWYPEIYVQALMGAGVEFLVSPPTPLMPSPPTVANIILGYDVGDKINSTCTWML 430
OY 364 R-----EETASPTPNATKPSPOISDSDSD 388
DB 431 KRYILGIEEPSPGIADVADKLKINSTD 462

RESULT 8
GUNE_NEOPA
ID GUNE_NEOPA STANDARD: PRT; 473 AA.
AC Q12647;
DT 01-NOV-1997 (rel. 35, Created)
DT 01-NOV-1997 (rel. 35, Last sequence update)

```

QY 61 TWQVYLVNTVDGKLDLWNGSYNKNVNAACLETGAYCMIDMNFARYNGGIIIGGCVSDDI 120  
 DB 166 RWE-RLQPTLVQVDFDANLSTRTGFAVAVTGTOTVLLDPHAYRYGNAVYSAVPSA 224  
 QY 121 FVDLWVOYAKYEDNDKRIEGLMNEPHDLDEIWAQTCQKVTARKGATISOMILLPGT 180  
 DB 225 YADWRRLATOFKSNPRVILGLMNEPNSMPTQWMLSGANAEILAIRSANA-SNVVFPVGN 283  
 QY 181 NFASV---ETVYSTGSAEALGKITPDGSTDLYFDVHKYLDIINSGSHAECTTDVDA 236  
 DB 284 AMTGAHSMNQWVYCTPRGTAVKGLNDPGHN---LVFEVHQLDSDSGSANCYSATIGA 340  
 QY 237 --FNDPFDMLRQNRKQAIISGTGASMEPSCMTAFCAONKAISENSDYVIGFVGAGSFD 294  
 DB 341 QRLDFTTTLRNSNGYRGFLGEFGAASNDTCQAVSNMILTEFKVNNADVWTGMWAGGPMW 400  
 QY 295 TSYLTLPGLKPGNYTDNKLNM 317  
 DB 401 GGYWYST---EPNGVYDKPQMS 419

## RESULT 5

GUN3\_HUMIN STANDARD; PRT; 388 AA.  
 AC Q12624; Q12620;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Endoglucanase 3 precursor (EC 3.2.1.4) (Endo-1,4-beta-glucanase 3) (cellulase 3).  
 GN CMC3.  
 OS Humicola insolens.  
 OC Eukaryota; Fungi; Ascomycota; mitosporic Ascomycota; Humicola.  
 OX NCBI\_TaxID=34413;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=94247364; PubMed=8190078;  
 RA Dalboege H., Hansen H.P.H.;  
 RT A novel method for efficient expression cloning of fungal enzyme genes.";  
 RT Mol. Genet. 243:253-260(1994).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=IFO 9854;  
 RA Takashima S., Nakamura A., Masaki H., Uozumi T.;  
 RT Cloning, sequencing and expression of the thermostable cellulase gene of Humicola grisea.";  
 RT Submitted (APR-1996) to the EMBL/GenBank/DBJ databases.  
 CC -1- CATALYTIC ACTIVITY: Endohydrolisis of 1,4-beta-D-glucosidic linkages in cellulose.  
 CC BIOCHEMISTRY: Used as a detergent cellulase. Sold under the name cellulzyme by Novozymes. This special enzyme has three effects: colour brightening, softening and removal of particulate soil. The overall effect is that it helps to preserve the nice appearance of new fabric and restores old fabric so that it looks new again.  
 CC -1- SIMILARITY: BELONGS TO CELLULOSE FAMILY A (FAMILY 5 OF GLYCOSYL HYDROLASES).  
 CC -1- SIMILARITY: CONTAINS 1 FUNGAL-TYPE CELLULOSE-BINDING DOMAIN (CBD).  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation- the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC EMBL: X76046; CAA53631.1;  
 DR EMBL: D84470; BAA12676.1;  
 DR HSSP: P00725; IAZ6.  
 DR InterPro: IPR000254; CBD\_fungal.  
 DR InterPro: IPR001547; GH\_5.  
 DR Pfam: PF00150; cellulase; 1.

DR Pfam: PF00734; CBM\_1; 1.  
 DR ProDom: PD001821; CBD\_fungal; 1.  
 DR SMART: SM00236; FCBP; 1.  
 DR PROSITE: PS00562; GLYCOSYL\_CBD\_FUNGAL; 1.  
 DR PROSITE: PS00659; GLYCOSYL\_HYDROL\_F5; FALSE\_NEG.  
 KW Cellulose degradation; Hydrolase; Glycosidase; Glycoprotein; Signal.  
 FT SIGNAL 1 16  
 FT CHAIN 17 388  
 FT DOMAIN 18 52  
 FT DOMAIN 53 91  
 FT DOMAIN 92 388  
 FT DISULFID 24 41  
 FT DISULFID 35 51  
 FT CARBOHYD 92 92  
 FT CARBOHYD 155 155  
 FT CARBOHYD 259 259  
 FT ACT\_SITE 215 215  
 FT ACT\_SITE 322 322  
 FT CONFLICT 8 8  
 FT CONFLICT 340 340  
 SO SEQUENCE 388 AA; 42563 MW; C7CF349DACCI0690 CRC64;

Query Match 14.58; Score 322; DB 1; Length 388;  
 Best Local Similarity 26.68; Pred. No. 3.9e-16;  
 Matches 88; Conservative 54; Mismatches 151; Indels 38; Gaps 8;

QY 1 GKIVYLVAVIPGIDFGCDIDGSCP-----TDTSSVPLLSYKGGDGAGCMKHPAEDD 51  
 DB 84 GKFWFGLNGSCAFG---KEYFGLMGKHTFPSTSI-----QTHI--ND 125  
 QY 52 GLAVFRISATWQVYLVNTVDGKLDLWNGSYNKNVNAACLETGAYCMIDMNFARYNGGII 111  
 DB 126 GFNFRVAFSMERLAPNQLNAPDANYLRNLTENVFTGKAYAMLDPHNFGRYERII 185  
 QY 112 GGGVSDDIFVDLWVOYAKYEDNDKRIEGLMNEPHDLDEIWAQTCQKVTARKGAT 171  
 DB 186 ---TDKAAFPSPFTKLTAFASNPVLEVDNNEHDMDDQLVDELQAAIDALIRAGAT 241  
 QY 172 SQMLLEPGTNFASVETVYSTGSAEALGKITPDGSTDLYFDVHKYLDIINSGSHAECTT 231  
 DB 242 SQYLMVEGNSWTCAMTWNVTNN--NLALRDEPK---LVQOMQYLDSDSGSTACVS 296  
 QY 232 DNV---DAFNDPFDMLRQNRKQAIISGTGASMEPSCMTAFCAONKAISENSDYVIGVWG 289  
 DB 297 TVQVLQRYIGATNMLRQNGKVLGEEFGAGANSVCOQIEGLMTHLDNSDVTGALWMA 356  
 QY 290 AGSPDTSYILTLPLKPGNTDKKLMECI 320  
 DB 357 GPPWMDYISFEPPSGIGYTYNSLKKYV 387

RESULT 6  
 GUN3\_ROBSP STANDARD; PRT; 375 AA.  
 AC P23044;  
 DT 01-NOV-1991 (Rel. 20, Created)  
 DT 15-JUN-2002 (Rel. 41, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Endoglucanase I (EC 3.2.1.4) (Endo-1,4-beta-glucanase) (Carboxymethyl-cellulase I) (CMCase I).  
 OS Robilliarida sp. (strain Y-20).  
 OC Eukaryota; Fungi; Ascomycota; mitosporic Ascomycota; Robilliarida.  
 OX NCBI\_TaxID=72589;  
 RN [1]  
 RP SEQUENCE OF 16-375 FROM N.A. AND SEQUENCE OF 1-18.  
 RX MEDLINE=91115784; PubMed=2277031;  
 RA Yoshigi N., Taniguchi H., Sasaki T.;  
 RT Cloning and sequencing of the endo-cellulase cDNA from Robilliarida sp. Y-20.";  
 RT J. Biochem. 108:388-392(1990).  
 CC -1- FUNCTION: ACTIVE TOWARDS CARBOXYMETHYL CELLULOSE.  
 CC -1- CATALYTIC ACTIVITY: Endohydrolisis of 1,4-beta-D-glucosidic linkages in cellulose.

```

RP SEQUENCE FROM N.A.
RC STRAIN-GM1100;
RX MEDLINE-21681879; PubMed-11823852;
RA Salanoubat M., Genin S., Artiguenave F., Gouzy J., Mangenot S.,
RA Ariat M., Billaut A., Brottier P., Camus J.C., Catolico L.,
RA Chaudier M., Choisme N., Claudel-Renard C., Cunac S., Demange N.,
RA Gaspin C., Layle M., Molan A., Robert C., Sautin W., Schlex T.,
RA Siguler P., Thebaud P., Whalen M., Winkler P., Levy M.,
RA Weissbach J., Boucher C.A.;
RT "Genome sequence of the plant pathogen Ralstonia solanacearum.";
RL Nature 415:497-502(2002).
CC -1- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic
CC linkages in cellulose.
CC -1- SUBCELLULAR LOCATION: Attached to the membrane by a lipid anchor
CC (Probable).
CC -1- SIMILARITY: BELONGS TO CELLULASE FAMILY A (FAMILY 5 OF GLYCOSYL
CC HYDROLASES).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL: AL646076; CAD17313.1; ALT_INIT.
DR InterPro: IPR001547; GH_5.
DR PROSITE: PS00013; PROKAR.LIPORPROTEIN: 1.
DR PROSITE: PS00659; GLYCOSYL_HYDROL_F5: 1.
KW Cellulose degradation; Hydrolase; Glycosidase; Zymogen; Membrane;
KW Lipoprotein; Signal; Complete proteome.
FT SIGNAL 1 19
FT PROPEP 20 43 BY SIMILARITY.
FT CHAIN 44 424 ENDOGLUCANASE.
FT LIPID 20 20
FT ACT_SITE 247 247 N-ACYL DIGLYCERIDE (BY SIMILARITY).
FT ACT_SITE 359 359 PROTON DONOR (BY SIMILARITY).
FT ACT_SITE 359 359 NUCLEOPHILE (BY SIMILARITY).
SQ SEQUENCE 424 AA; 45080 MW; 5BF0D161AB7C87EF CRC64;

Query Match 18.1%; Score 400; DB 1; Length 424;
Best Local Similarity 29.4%; Pred. No. 1e-21;
Matches 93; Conservative 60; Mismatches 14; Indels 24; Gaps 9;

QY 1 GKIKYLVAIPIGIDGCDIDGSCPTDTSSVPLSLYKGGAGAGOMKHFADDELNFRISA 60
DB 113 GTLLMRGVSLAGAEFG---EGSLPGTYSNYT--YPSADSVTYRK---NKGMLVRLPF 163
QY 61 TWQFVLTNTVDGKDELMLMGSTNRYVNAACLETGAYCMIDMNFARYNGIIGOGVSDDI 120
DB 164 RWE-RIQFTLNQALDANELSLRTLGFTVNAVTAAGQVLLDPHYARYRYGVSSAIPNSA 222
QY 121 FVDLWVQIAKYEDNDKIIFGLMNEPHDIEIMVQTOCKVYTAIRKAGATGOMILLPRT 180
DB 223 YADFFRRVATOKGNARVYIFGLMNEPNSPTQOMLSGANAALAAIRSAVA-SNVVFPNGN 281
QY 181 NEASY-----ETVYSTGSAEALGKITNPDSOTDLFYDVYKRYDILNNGSHACTTDNVA 236
DB 282 AWTGAHSMNQNMVYGTPTNGVYMGINDPGRN---LVFEVHYQYIDGSSSGSACYSATICA 338
QY 237 --FNDPADWLNRONKROAIITSETGASMEPSCMATFCAONKAISNSVDYVIGYVGAGSFD 294
DB 339 ERLDFTWMLNSNGYRGELGFEFGAASNDTCQAVANMLTFEVNNDVDTGMAMWAGGPMW 398
QY 295 TSYITLTLPPLKPGNYTDNKLKN 317
DB 399 GGYMYST---EPSNGVDPKPM 417

```

```

DT 01-NOV-1990 (Rel. 16, Created)
DT 01-MAY-1992 (Rel. 22, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Endoglucanase precursor (EC 3.2.1.4) (Endo-1,4-beta-glucanase)
DE (Cellulase).
GN Egl.
OS Ralstonia solanacearum (Pseudomonas solanacearum).
OC Bacteria; Proteobacteria; beta subdivision; Ralstonia group;
OC Ralstonia.
OX NCBI_TaxID=305;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-AW;
RX MEDLINE-92138626; PubMed-1735723;
RA Huang J., Schell M.A.;
RT "Role of the two-component leader sequence and mature amino acid
RT sequences in extracellular export of endoglucanase Egl from
RT Pseudomonas solanacearum.";
RL J. Bacteriol. 174:1314-1323(1992).
RN [2]
RP SEQUENCE OF 1-112 FROM N.A., AND PARTIAL SEQUENCE.
RX MEDLINE-89291722; PubMed-2738021;
RA Huang J., Sukoradhan M., Schell M.A.;
RT "Excretion of the egl gene product of Pseudomonas solanacearum.";
RL J. Bacteriol. 171:3767-3774(1989).
RN [3]
RP PROCESSING.
RX MEDLINE-90307678; PubMed-2195024;
RA Huang J., Schell M.A.;
RT "Evidence that extracellular export of the endoglucanase encoded by
RT egl of Pseudomonas solanacearum occurs by a two-step process
RT involving a lipoprotein intermediate.";
RL J. Biol. Chem. 265:11628-11632(1990).
CC -1- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic
CC linkages in cellulose.
CC -1- SUBCELLULAR LOCATION: Attached to the membrane by a lipid anchor
CC (Probable).
CC -1- SIMILARITY: BELONGS TO CELLULASE FAMILY A (FAMILY 5 OF GLYCOSYL
CC HYDROLASES).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL: M84922; AA61980.1; -.
DR PIR: A32884; A32884.
DR PIR: A42649; A42649.
DR InterPro: IPR001547; GH_5.
DR Pfam: PF00150; Cellulase: 1.
DR PROSITE: PS00013; PROKAR.LIPORPROTEIN: 1.
DR PROSITE: PS00659; GLYCOSYL_HYDROL_F5: 1.
KW Cellulose degradation; Hydrolase; Glycosidase; Zymogen; Membrane;
KW Lipoprotein; Signal.
FT SIGNAL 1 19
FT PROPEP 20 45
FT CHAIN 46 426 ENDOGLUCANASE.
FT LIPID 20 20
FT ACT_SITE 249 249 N-ACYL DIGLYCERIDE.
FT ACT_SITE 361 361 PROTON DONOR (BY SIMILARITY).
FT ACT_SITE 361 361 NUCLEOPHILE (BY SIMILARITY).
SQ SEQUENCE 426 AA; 45578 MW; 51E13AD4442C4A8 CRC64;

Query Match 17.4%; Score 386; DB 1; Length 426;
Best Local Similarity 28.8%; Pred. No. 1e-20;
Matches 93; Conservative 59; Mismatches 14; Indels 24; Gaps 9;

QY 1 GKIKYLVAIPIGIDGCDIDGSCPTDTSSVPLSLYKGGAGAGOMKHFADDELNFRISA 60
DB 115 GTLLMRGVSLAGAEFG---EGSLPGTYSNYT--YPSADSVTYRK---NKGMLVRLPF 165

```

DR InterPro: IPR000254; CBD\_fungal.  
 DR InterPro: IPR001547; GH\_5.  
 DR Pfam: PF00150; cellulase; 1.  
 DR Pfam: PF00734; CBM\_1; 1.  
 DR ProDom: PD001821; CBD\_fungal; 1.  
 DR SMART: SM00336; fCBD; 1.  
 DR PROSITE: PS00562; CBD\_FUNGAL; 1.  
 DR PROSITE: PS00659; GLYCOSYL\_HYDROL\_F5; 1.  
 DR Cellulose degradation; Hydrolase; Glycosidase; Glycoprotein; Signal.  
 FT SIGNAL 1 21  
 FT CHAIN 22 418  
 FT DOMAIN 22 57  
 FT DOMAIN 58 91  
 FT MOD\_RES 92 418  
 FT MOD\_RES 22 22  
 FT CARBOHYD 124 124  
 FT DISULFID 29 46  
 FT DISULFID 40 56  
 FT ACT\_SITE 239 239  
 FT ACT\_SITE 350 350  
 FT ACT\_SITE 350 350  
 SQ SEQUENCE 418 AA: 44227 MW: 26449255237A49B CRC64;

Query Match 42.6%; Score 942.5; DB 1; Length 418;  
 Best Local Similarity 51.7%; Pred. No. 6.2e-61;  
 Matches 167; Conservative 61; Mismatches 90; Indels 5; Gaps 2;

QY 3 IKYLVAIPGIDFGCDIDSCPTDTSVPLSYKRG---DGAGQMKHFAEDGLNFR 58  
 DB 93 VRFAGVNIAGDFGCTDTCGTCVTSKYVPLKNTGSGNNYPRGICOMGFVDEDMITFRL 152  
 QY 59 SATWQFVNLNTVDGKLDLNMGSYKYNVACLETGAYCMIDMNFARNGGIIIGGGVSD 118  
 DB 153 PVGQOYLVNNLNGLDSTSIKRYDQGLGAYCIYDINHVARNGGIIIGGGPTN 212  
 QY 119 DIFDVLVQIAKYEDNDKIIFGLMNEPHDLIEYNAQTOCKYVTAIRKAGATSQMILP 178  
 DB 213 AQFSLMSQLASKTASQSRWFGIMNEPHDNTVMAATVOEVYTAIRNAGATSQFISLP 272  
 QY 179 GTNFASVETVSTGSALGKITNPDSSTDLVDFVHKYLDINNSGSHACTDNDV-AF 237  
 DB 273 GNDQSGAFIDSGAALSGVTNPDSSTDLVDFVHKYLDINNSGSHACTDNDV-AF 332  
 QY 238 NDFADMLRQNRKQAIISSETGASMEPSCMTAFCAONKAISENSDVYIGFVGAGSFDTSY 297  
 DB 333 SPLATWLRQNRKQAIILETETGCGVNOQCIODMCQIOYLNQNSDVYIGFVGAGSFDTSY 392  
 QY 298 ILTLPKPGKPNYDNKLMNECI 320  
 DB 393 VLETPTSSGNSWTDTSVSSCL 415

RESULT 2  
 GUN1\_CRYFL STANDARD; PRT; 341 AA.

AC 004469;  
 DT 01-OCT-1996 (Rel. 34, Created)  
 DT 01-OCT-1996 (Rel. 34, Last sequence update)  
 DT 30-MAY-2000 (Rel. 39, Last annotation update)  
 DE Endoglucanase 1 precursor (EC 3.2.1.4) (Endo-1,4-beta-glucanase 1)  
 DE (Carboxymethyl-cellulase 1) (CMCase 1) (Cellulase 1).  
 GN CMCL.  
 OS Cryptococcus flauus.  
 OC Eukaryota; Fungi; Basidiomycota; Hymenomycetes; Heterobasidiomycetes;  
 OC Tremellomycetidae; Tremellales; mitosporic Tremellales; Cryptococcus.  
 OX NCBI\_TaxID=5416;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=93005075; PubMed=1368837;  
 RA Cui Z., Mochizuki D., Matsuno Y., Nakamura T., Liu Y., Hatano T.,  
 RA Fukui S., Miyakawa T.,  
 RT Cloning and molecular analysis of cDNA encoding a  
 RT carboxymethylcellulase of the yeast Cryptococcus flauus.  
 RL Biosci. Biotechnol. Biochem. 56:1230-1235(1992).

CC -1- FUNCTION: HAS ENDOGLUCANASE ACTIVITY ON CARBOXYMETHYL-CELLULOSE  
 CC (CMC).  
 CC -1- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic  
 CC linkages in cellulose.  
 CC -1- SIMILARITY: BELONGS TO CELLULOSE FAMILY A (FAMILY 5 OF GLYCOSYL  
 CC HYDROLASES).  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: D13967; AAC03070.1; -;  
 DR EMBL: S45137; BAC60541.1; -;  
 DR InterPro: IPR001547; GH\_5.  
 DR Pfam: PF00150; cellulase; 1.  
 DR PROSITE: PS00659; GLYCOSYL\_HYDROL\_F5; 1.  
 DR Cellulose degradation; Hydrolase; Glycosidase; Signal.  
 FT SIGNAL 1 16  
 FT PROPEP 17 30  
 FT CHAIN 31 341  
 FT ACT\_SITE 166 166  
 FT ACT\_SITE 275 275  
 FT ACT\_SITE 275 275  
 SQ SEQUENCE 341 AA: 35717 MW: 4A62E45C952DD120 CRC64;

Query Match 29.5%; Score 654; DB 1; Length 341;  
 Best Local Similarity 44.4%; Pred. No. 3.3e-40;  
 Matches 139; Conservative 47; Mismatches 109; Indels 18; Gaps 7;

QY 7 GVALPGIDFGDIDSCPTDTSVPLSYKRGDAGQMKHFAEDGLNFRISATMGOVL 66  
 DB 37 GVALAGCDFGIDTIGN--SGTPACP-----GTBOVGHPIA-DGANLFLPAGMOYLV 85  
 QY 67 -NNTVDKRLDELNMGSYKYNVACLETGAYCMIDMNFARNGGIIIGGGVSDIFVDLW 125  
 DB 86 GNNQASTSLADPFAQYDALVOAVISKGAYALIDVHNHVARNGGIIIGGGSPNDPFLNW 145  
 QY 126 VQIAKYEDND-KIIFGLMNEPHDLIEYNAQTOCKYVTAIRKAGATSQMILPCTNFAS 184  
 DB 146 TLATKVTYSDNPVIFGLMNEPHDLVSTVAGVQAAVNAIRAGATSQYLLIGTGFETN 205  
 QY 185 VETVSTGSALGKITNPDSSTDLVDFVHKYLDINNSGSHACTDNDVAFNPDWLT 244  
 DB 206 ANAMFQ-GQDNALLAGVDVPGGTGKLLDVHRYNDVPSGTHACTTNSLDVSSLSWL 264  
 QY 245 RQNRKQAIISSETGASMEPSCMTAFCAONKAISENSDVYIGFVGAGSFDTSYILTLPL 304  
 DB 265 KGNCRKAIVSEFGGHTTSCETDGLGEPLNGIKEDYPSVLGFAVMAAGSFDTSYLSITPT 324  
 QY 305 GKPGNYDNKLMN 317  
 DB 325 ----NGVDNQLFD 333

RESULT 3

GUN1\_RALSO STANDARD; PRT; 424 AA.

AC P58599;  
 DT 15-JUN-2002 (Rel. 41, Created)  
 DT 15-JUN-2002 (Rel. 41, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Endoglucanase precursor (EC 3.2.1.4) (Endo-1,4-beta-glucanase)  
 DE (cellulase).  
 GN EGL OR RSP0162 OR RS05516.  
 OS Ralstonia solanacearum (Pseudomonas solanacearum).  
 OC Plasmid megaplasmid.  
 OC Bacteria; Proteobacteria; beta subdivision; Ralstonia group;  
 OC Ralstonia.  
 OX NCBI\_TaxID=305;  
 RN [1]



**THIS PAGE BLANK (USPTO)**

```

Db      204 NNPSNDIREAVACINDYNOVALDAIRATGNNATRCVWPGYD-ASIEGCMTDGF----- 257
OY      199 KITNPDGSDLLYFDYHKYLDIN---NSGSHACTDN---VDA--FNDEADWLRONKRQ 250
Db      258 KRN-DTASGRLLSYHATIPYFALASDYYTRFDNKLKYDIDSEFNLSKFLSRNP 316
OY      251 AIISGTASMEPSCMTAFCAONKAISEN-SDVYIGFVGWGSFDTSYLITPLGKPGN 309
Db      317 VVVGESATNR-----NNTAEVRKNADY-----WGRARYSNVAMVL----- 354
OY      310 YTDNKLKLM-----NEC--ILDQFTLDEKRYPTPTISITAEETATATSDG-DAPST 358
Db      355 -MDNNTYQNNNSAGSDECHMYIDRNSLOWK---DPEIISTYMKHVDGTPATINGKEIPST 410
OY      359 TKPIFREETASPTPNAVTKPSPTSDSDDDKD 391
Db      411 EOPDPTPVDDPTP-VDPDPTPVDDPTPVDPD 442

```

## RESULT 15

S16559

cellulase (EC 3.2.1.4) - Ruminococcus flavefaciens

C:Species: Ruminococcus flavefaciens

C:Date: 19-Mar-1997 #sequence\_revision 19-Mar-1997 #ext\_change 24-Nov-1999

C:Accession: S16559

R.Cunningham, C.; McPherson, C.A.; Martin, J.; Harris, W.J.; Flint, H.J.

Mol. Gen. Genet. 228, 320-323, 1991

A:Title: Sequence of a cellulase gene from the rumen anaerobe Ruminococcus flavefaciens

A:Reference number: S16559; MUID:91360084; PMID:1886616

A:Accession: S16559

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-455 &lt;CUN&gt;

A:Cross-references: GB:S55178; NID:9234871; PIDN:AAB19708.1; PID:9234872

C:Superfamily: cellulase CCA; Clostridium cellulase; repeat homology

C:Keywords: glycosidase; hydrolase

## Query Match

6.9%; Score 153; DB 2;

Best Local Similarity: 24.7%; Pred. No. 0.0018; Length 455;

Matches 60; Conservative 41; Mismatches 92; Indels 50; Gaps 12;

```

OY      52 GLNFRISATWQVLNNTYDGLDELNMGSYNKNVACLETSAVCMIDMHNPARYNGI 111
Db      102 GFNTVRIPTTWQHLDE--NDNIDPAMMARVHQVVDYAVNIILYVILNLHQQNW---I 155
OY      112 GGGGVS---DDI---FVDLWQIAKYEDNDK-IIFGLNRPDLDI--ETAAQTCQKY 162
Db      156 NRADLATAYDDINPRLMKLTQIATEFKDYDQHLIFECNNEPRAMDTPWEMWSATPVEER 215
OY      163 TAIRKAGAT-----SOMILPGTNFASVETVSTGSAELCKITNPDGSTD 208
Db      216 DVINRLEAFNVELIRMGDPYAKTRILMPG-----YVASSDKTFILQIYLPE-ND 266
OY      209 LLYFDVHKYLDIN-----NSGSHAECTT---DNVDAFN--DFADILRONKROAISET 256
Db      267 FLAVSIHAYTPYNTFTMTTEGAYHDFTFKERSNDLAYNLQNFQDMFINKDIPVIGEM 326
OY      257 GAS 259
Db      327 GTS 329

```

Search completed: May 9, 2003, 15:21:24  
 Job time : 30.742 secs

Db 735 ALQGRGLEVVAVNAVEKKPEEPKTEPEPTTETTSPEEPTETTNPEPTGNIRDISK 794  
 QY 388 -----DKRDSASMSAOG 400  
 Db 795 ELIKEMNFGNMLNTLDAQCEIYLNKDKQDTASFTCWG 832

## RESULT 12

S40507

endoglucanase - rumen fungus (Neocallimastix patriciarum)

C:Species: Neocallimastix patriciarum

C&gt;Date: 19-Mar-1997 #sequence\_revision 19-Mar-1997 #text\_change 31-Mar-2000

C:Accession: S40507

R:Zhou, L.; Xue, G.; Orpin, C.G.; Black, G.W.; Gilbert, H.J.; Hazlewood, G.P.

Biochem. J. 297, 359-364, 1994

A:Title: Intracellular cels from the anaerobic fungus Neocallimastix patriciarum encodes a

A:Reference number: S40507; PMID:94128068; PMID:8297343

A:Accession: S40507

A&gt;Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-473 &lt;HO&gt;

A:Cross-references: GB:231364; GB:X77186; NID:9467686; PIDN:CA63238.1; PID:9467687

C:Superfamily: rumen fungus cellulase; Clostridium cellulase repeat homology

Query Match 7.98; Score 175.5; DB 2; Length 473;

Best Local Similarity 22.28; Pred. No. 4.4e-05;

Matches 88; Conservative 55; Mismatches 172; Indels 81; Gaps 17;

QY 12 GIDGCDIDGSCPTDTSVPLLSYKGGDGA-----GQKHFRAE-----DGLNVRISA 60  
 Db 35 GWSLGNITLDAIC-----FETLDYNNKQIASSETGKNTTOLLYKSLDGLGFTFRLPT 88  
 QY 61 TWQFVLANNTVDGKIDELNMGSYNKNVNACTETGAYCMIDMNFARNGIIGOGVSDDI 120  
 Db 89 TWSGHFENADPYKINDQMKRHEIVDYALINTGYAILNLH-ETWNAFQKNLESARKI 147  
 QY 121 FVDLMVQIAKYEDNDK-IIFGLNPERDLDIET-MAOTCK-----VVTAIRK 167  
 Db 148 LVAITMKQIAEFADYDEHLIFEGMNEPRKYDPAEMNGDDEGNFVEMNDLFEVKTIRA 207  
 QY 168 AGATSQM-ILLPGTNFASVETVSTGSAAELAGKITNPDGSTDLLFEDVHKYLDIN--N 222  
 Db 208 TGGNNALRHLMIIP-----TYAACINDGAINNFKPSSG-DKVIYLSHSYSTNFRALN 258  
 QY 223 SCSHAECTTDVDAFNDA---DWLRQNRQAIISETGASMEPSCMTAFCAONKAISENSD 280  
 Db 259 NGAGA-----ISNFYDGEIDMAMNTINSKFSRG-----LPIVIGERGAMNRNNDRE 308  
 QY 281 VYIGFVGWAGSFPDTSYILITPLKRP-----GNTDNKMLNCEILDQFTIDEXY----- 330  
 Db 309 RMAEY-----YIKKATISIGVPCVIWDNCGYFEGEGEFGLINRSTLQVVYPRKLVN 357  
 QY 331 ---RPTPTISATAETATATSDGAPSTTRPIF 363  
 Db 358 GLIKGLGNSIKTRITRTTITTTTTSQSOPNTNDSCF 393

## RESULT 13

JE0302

cellulase (EC 3.2.1.4) - rumen fungus (Neocallimastix frontalis)

N:Alternate names: endo-1,4-beta-glucanase

C:Species: Neocallimastix frontalis

C&gt;Date: 03-Dec-1998 #sequence\_revision 05-Dec-1998 #text\_change 21-Jul-2000

C:Accession: JE0302

R:Fujino, Y.; Ogata, K.; Nagamine, T.; Ushida, K.

Biosci. Biotechnol. Biochem. 62, 1795-1798, 1998

A:Title: Cloning, sequencing, and expression of an endoglucanase gene from the rumen and

A:Reference number: JE0302; PMID:99022200; PMID:9805384

A:Accession: JE0302

A:Molecule type: mRNA

A:Residues: 1-482 &lt;FUJ&gt;

A:Cross-references: GB:U38843; NID:g1051284; PIDN:AAC63094.1; PID:g3712668

C:Genetics:

A:Gene: cels  
 C:Superfamily: endoglucanase; bacterial cellulose-binding domain homology  
 C:Keywords: glycosidase; hydrolase

Query Match 7.48; Score 163.5; DB 2; Length 482;

Best Local Similarity 20.38; Pred. No. 0.00034;

Matches 84; Conservative 70; Mismatches 164; Indels 95; Gaps 18;

QY 12 GIDGCDIDGSCPTDTSVPLLSYKGGDGA-----GQKHFRAE-----DGLNVRISA 60  
 Db 53 GWSLGNITLDAIC-----FETLDYNNKQIASSETGKNTTOLLYKSLDGLGFTFRLPT 106  
 QY 61 TWQFVLANNTVDGKIDELNMGSYNKNVNACTETGAYCMIDMNFARNGIIGOGVSDDI 120  
 Db 107 TWSGHFENADPYKINEKRLRHEIVDYPRKNGAFVILNLH-ETWNAFSETLDTAKEI 165  
 QY 121 FVDLMVQIAKYEDNDK-IIFGLNPERDLDIET-MAOTCK-----VVTAIRK 167  
 Db 166 LEKIMSQIAKEFKYDEHLIFGLNPERKNDTPVEMTGGDGDGDVAVNNAVELKTRIS 225  
 QY 168 AGATS---OMILLPGTNFASVETVSTGSAAELAGKITNPDGSTDLLFEDVHKYLDIN--N 222  
 Db 226 SGNPNRHLMIIP-----PYAACNENSPKNIPIFE-DDDKVIASVHAHAYARNFRALN 276  
 QY 223 SCSHAECTTDVDAFNDA---DWLRQNRQAIISETGASMEPSCMTAFCAONKAISEN 278  
 Db 277 NGAGA-----VDRFDAGKKDLNINLMKKRFVDCG---IPMILGEYCAMNRNDNEE 326  
 QY 279 SDVIYIGFVGWAGSFPDTSYILITPLKRPQVYTNKMLNCE---LIDOTLDEKRYPT 333  
 Db 327 RATAEY-----YMEKVTAMGVQVWMDNGVEGTGERGLDRLNLITVY--- 372  
 QY 334 PTISITA-----AETATATATSDG---DAPSTTRPIFEETAS 369  
 Db 373 PTIAALQKGRGLEVVAVNAVEETEBECMSKEYECCSPNNTVYVSDSGN 425

## RESULT 14

J00356

cellulase (EC 3.2.1.4) - Butyrivibrio fibrisolvens

N:Alternate names: endo-1,4-beta-glucanase

C:Species: Butyrivibrio fibrisolvens

C&gt;Date: 07-Sep-1990 #sequence\_revision 07-Sep-1990 #text\_change 08-Oct-1999

C:Accession: J00356

R:Berger, E.; Jones, W.A.; Jones, D.T.; Woods, D.R.

Mol. Gen. Genet. 219, 193-198, 1989

A:Title: Cloning and sequencing of an endoglucanase (end1) gene from Butyrivibrio fib

A:Reference number: J00356; PMID:90136507; PMID:2615759

A:Accession: J00356

A:Molecule type: DNA

A:Residues: 1-547 &lt;BER&gt;

A:Cross-references: GB:X17538; NID:g39472; PIDN:CA55574.1; PID:g39473

A:Experimental sources: strain H17C

C:Comment: B. fibrisolvens is an important proteolytic, cellulolytic, and hemocellulo

C:Function: hydrolysis of 1,4-beta-D-glucosidic linkages in beta-D-glucans such as

A:Pathway: cellulose degradation

C:Superfamily: endoglucanase; bacterial cellulose-binding domain homology

C:Keywords: glycosidase; hydrolase; polysaccharide degradation

Query Match 7.28; Score 158.5; DB 2; Length 547;

Best Local Similarity 23.74; Pred. No. 0.00093;

Matches 93; Conservative 56; Mismatches 149; Indels 95; Gaps 25;

QY 52 GLNVRISATW-QPVLANNTVDGKIDELNMGSYNKNVNACTETGAYCMIDMNF-----F 103  
 Db 92 GFNTIRIPVSGQYTTGS--DYQIPDVYMNKVKAEVVDCTIYNDKVIYLSNHHINDSYCF 149  
 QY 104 ARYNGIIGOGVSDDI-FVDLMVQIAKYEDND-KIIFGLNPERDLDIETIY----- 154  
 Db 150 YVPRNN---ANKDRSEKFKFSWQIAKEFKNYDVHLVETINMERLVLGHG---EDMNFPR 203  
 QY 155 -----AOTC-----QKVYTAIRKAGA--TSOMILLPGTNFASVETVSTGSAAELG 198

F:76,182,219,271/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 9.8% Score 217.5; DB 2; Length 375;

Best Local Similarity 24.9%; Pred. No. 2,9e-08; Matches 82; Conservative 50; Mismatches 128; Indels 69; Gaps 14;

OY 3 IKYLYAIPGIDPCDIDSCPTDTSVPLSYKGDGA--GOMKHEADDGLANFRI 58  
 2 IYVAGVAGGEGFV---WSATQTPGTGLPGRFVGYANISAAVDVHDOHNLNFRV 57  
 OY 59 SATMOVNLNT-----VDKIDELMWGSYNKYVNAQLEF-GAYCMIIDHNFARYN--- 107  
 58 A---FLIERMCPKPAAGLGAANFETHEFKEAVDYITVYKGAVALIDPHNMYKYNLAT 113  
 OY 108 -----GIIIGGGVSDIFVYDLW-----VOIARYEDNDKIRFGLMN 144  
 DB 114 SLGREVSLATPQTPQPPRSLASFVS-----WHLASMTSHAFHMAHERKTRIG--- 163  
 OY 145 EPHDDIEIWAOTQKRVYTAIRKAGATSQMILLPTNFASVETV--STGSABALGKITN 202  
 DB 164 -----DLVVLANN-QAIDAIRANA-SNLIIMPNSWGTGHSWTEGSDPSALLNQFD 216  
 OY 203 POSTDLLEFDVHKYLDINSSGHACTDNDVAFNDFADWLQONKROAISETGASMER 262  
 DB 217 PLNNIAI---DIHEYLDYDPSGGHLECVSDPETNLAALTAMLEKNLKAITEFGGSNST 273  
 OY 263 SCMTAFCAONKAISENSDYVYGFVGWAG 291  
 DB 274 SCOEMLPDLINYMADNAE-YIGWTANAG 301

# RESULT 10

CELLULASE (EC 3.2.1.4) E precursor - Clostridium thermocellum  
 N:Alternate names: endo-1,4-beta-glucanase E; endoglucanase E  
 C:Species: Clostridium thermocellum  
 C>Date: 31-Mar-1992 #sequence\_revision 31-Mar-1992 #text\_change 18-Jun-1999  
 C:Accession: J0347  
 R:Hall, J.; Hazlewood, G.P.; Barker, P.J.; Gilbert, H.J.  
 Gene 69, 29-38, 1988  
 A:Title: Conserved reiterated domains in Clostridium thermocellum endoglucanases are not  
 A:Reference number: J0347; M01D:89137992; PMID:3066598  
 A:Accession: J0347  
 A:Molecule type: DNA  
 A:Residues: 1-814 <HAL>  
 A:Cross-references: GB:M22759; NID:q144768; PIDN:AA23224.1; PID:q144770  
 A:Note: part of this sequence, including the amino end of the mature protein, was confir  
 C:Comment: This secretory enzyme is part of a highly active and thermostable cellulase c  
 C:Genetics:  
 A:Gene: cele  
 A:Start codon: GTG  
 C:Function:  
 A:Description: catalyzes the hydrolysis of 1,4-beta-D-glucosidic bonds in beta-D-glucans  
 A:Pathway: cellulose degradation  
 C:Superfamily: cellulase CCA; Clostridium cellulase repeat homology  
 C:Keywords: duplication; extracellular protein; glycosidase; hydrolase; polysaccharide c  
 F:1-34/Domain: signal sequence #status predicted <SIG>  
 F:35-814/Product: cellulase E #status predicted <MAT>  
 F:415-438/Domain: Clostridium cellulase repeat homology <CCN1>  
 F:451-474/Domain: Clostridium cellulase repeat homology <CCR2>

Query Match 8.7% Score 193.5; DB 1; Length 814;

Best Local Similarity 23.5%; Pred. No. 4,6e-06; Matches 92; Conservative 60; Mismatches 157; Indels 83; Gaps 19;

OY 52 GINVERISATWQFVLNNTVDGKLDELNMGSYNKNVNAQLETAAYCMIDMHNFARYNGII 111  
 DB 99 GRNAVAVPTWPTDHIGRPADYKIDKIDAMLRNVEEVYVLDGMYAIILNHHDTWIIIPY 158  
 OY 112 GGGGVSDDIFVDMVOIAKYED-NDKIIFGLMNEPHDL--DIETWAQTCQK----- 160  
 DB 159 ANEORSKEKLIVWMOIATRFKDYDHLLEFETNNEPREVGSPEWMMGTYENRDVIINREN 218

OY 161 --VTAIRKAGATSQMILLPTNFASVETVSTGSALCKITNPDSGLLYEDVHKY- 217  
 DB 219 LAVVNTIRASGNN-----KRFILPTMAAGLDVALNDIVIPNDSVVI--VSHIAYS 271  
 OY 218 -----LDINSS--GSHAE--CTTDVDA-FNDPADLRONKROAISETGASMERSCAT 266  
 DB 272 PYEFAMDVNQTSGWSDYDKASLTSELDIAYNRFV-----KNGRAVIIGFEG-TIDKNLS 326  
 OY 267 AFCAQ-----NKAISENSDYVYGFVGAGSPFSTYLLTLPKPGNYNDKLMNLCIL 321  
 DB 327 SRVAHNEHVARERAVSGIAF-----W---WDNGTY-----NPDATYALLNKKTL 370  
 OY 322 DQF-----TLDKRYRPTPT-SISTAEEATAT-----ATSDGASTTKPIF 363  
 DB 371 SWYPEIYQALMGAGVEPLVSPPTPLMPTSPPTATILYGDVNGDKINSTDCTML 430  
 OY 364 R-----EETASPTPAVATKPSPTDSSDD 388  
 DB 431 KRYLRGIEEFPSPGIADVADLKINSTD 462

# RESULT 11

CELLULASE (EC 3.2.1.4) CelD - rumen fungus (Neocallimastix patriciarum) (fragment)  
 C:Species: Neocallimastix patriciarum  
 C>Date: 29-Oct-1999 #sequence\_revision 29-Oct-1999 #text\_change 29-Oct-1999  
 C:Accession: T31426  
 R:Alward, J.H.; Gobius, K.S.; Kennedy, P.M.; Simpson, G.D.; Xue, G.P.; Dalrymple, B  
 submitted to the EMBL Data Library, March 1998  
 A:Description: Characterisation of a Neocallimastix patriciarum cellulase CelD and co  
 A:Reference number: 221026  
 A:Accession: T31426  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: mRNA  
 A:Residues: 11232 <AYL>  
 A:Cross-references: EMBL:AF053363; NID:q2981483; PID:q2981484; PIDN:AAC06321.1  
 C:Genetics:  
 A:Gene: celD  
 C:Function:  
 A:Description: hydrolysis 1,4-beta-D-glucosidic linkages in cellulose  
 A:Pathway: cellulose degradation  
 C:Keywords: glycosidase; hydrolase

Query Match 8.4% Score 186; DB 2; Length 1232;

Best Local Similarity 20.5%; Pred. No. 2,9e-05; Matches 94; Conservative 69; Mismatches 191; Indels 104; Gaps 19;

OY 12 GIDFGCDIDGSCPTDTSVPLSYKGDGAGQ-----MKHFAEDGLNVERISA 60  
 DB 410 GWNIGNTLDAQ-----IEVLNVDKQDTASETCWGNPKTTEDMFKYLMNDFNFRIP 463  
 OY 61 TMOFVLNNTVDGKLDELNMGSYNKNVNAQLETAAYCMIDMHNFARYNGIIGGGVSDI 120  
 DB 464 TWGNGHGEADYVINEMLKRVHEIYDYPKNAFVILNHN-ETWNAHSETLDTAKEI 522  
 OY 121 FVDMVOIAKYEDNDK-IIFGLMNEPHDIEI-----W-AQTCQKV--TAIRK 167  
 DB 523 LKIMNSIAEERFDYDEHLEFEGLENERKNDTYESTGQOEGMDAVNANNAFLKTI 582  
 OY 168 AGATS--QMILLPTNFASVETVSTGSALCKITNPDSGLLYEDVHKYLDINSSGS 225  
 DB 583 SGNPNRRLHMLP-----PYAACNENSPKNTFEP-EDDKVILASVAVAPYFALN 633  
 OY 226 HACTTDNDVAF--NDPADLRONKROAISETGASMERSCATFCAONKAISENSVYT 263  
 DB 634 NGEAVDKFPAACKNDL-ENINIMMKRFVDOG---IPMILDEYGMNNDNEEDRAWA 688  
 OY 284 GFVGWAGSPFSTYLLTLPKPGNYTDKLMNEC-----IIDQFTLDEKYRPTPT 338  
 DB 689 EF-----YMEKVTAMGVPOYWMNDGITEGGERGLDRLRLKLYV---PTIYA 734  
 OY 339 TAAE-----ETATATATSGDAPSTTKPIFREETASP--PPNAVTPSPPTSSSD-- 387

OY 237 --FNFAFWLROMKQALIISETGASMEPSCMTAFCAONKKAISENDVYIGFVWGAGSEFD 294  
 Db 341 ORLODPTTWLWSNCRGLRGLGEFGAASNDTCNOAVSNMLTFYKNNADVNTGAMWAGGFW 400  
 OY 295 TSYILTLPLGKPGNYTDNKLIN 317  
 Db 401 GGMYSI-----EPSNGYDKPQWS 419  
 RESULT 7  
 JC5461  
 cellulase (EC 3.2.1.4) precursor - imperfect fungus (Humicola grisea)  
 N:Alternate names: endo-1,4-beta-glucanase; endoglucanase  
 C:Species: Humicola grisea var. thermoides  
 C:Date: 11-Sep-1998 #sequence\_revision 11-Sep-1998 #text\_change 16-Jun-2000  
 C:Accession: JC5461  
 R:Takashima, S.; Nakamura, A.; Masaki, H.; Uozumi, T.  
 Biosci. Biotechnol. Biochem. 61, 245-250, 1997  
 A:Title: Cloning, sequencing, and expression of a thermostable cellulase gene of Humicola  
 A:Reference number: JC5461; MUID:97212020; PMID:9058960  
 A:Accession: JC5461  
 A:Molecule type: DNA  
 A:Residues: 1-388 <TRAX>  
 A:Cross-references: DDBJ:D84470; NID:g1304101; PIDN:BAAL2676.1; PID:g1304102  
 A:Experimental source: strain IF09854  
 C:Genetics:  
 A:Gene: eg12  
 A:Introns: 120/3; 369/1  
 C:Function:  
 A:Description: hydrolysis of 1,4-beta-D-glucosidic linkages in beta-D-glucans such as cel  
 A:Pathway: cellulose degradation  
 C:Superfamily: Trichoderma cellulase III; fungal cellulose-binding domain homology  
 C:Keywords: glycosidase; hydrolase; polysaccharide degradation  
 F:1-16/Domain: signal sequence #status predicted <SIG>  
 F:17-388/Product: cellulase #status predicted <MAT>  
 F:21-52/Domain: fungal cellulose-binding domain homology <FCB>  
 Query Match 14.6%; Score 323; DB 1; Length 388;  
 Best Local Similarity 26.6%; Pred. No. 6.7e-16;  
 Matches 88; Conservative 54; Mismatches 151; Indels 38; Gaps 8;  
 OY 1 GKIKYLVALPIGIDFGCIDGSCP-----TDSSVPLLSYKGGDAGOMKHFADD 51  
 Db 84 GKFKMGINGOSCAEFG--KGEYPLGMCKHFFPSTSI-----QTHI--ND 125  
 OY 52 GLNVPRISATNOQFVLANTVDDKLDLNMGSYKVKYVNAACLENGAYCQIMHNFARKNGIT 111  
 Db 126 GNNRWFVAFSEMRILAPNOLNAAFADANYLRNLLETFTGKGKYMPLDPHNGRYERII 185  
 OY 112 GCGGVSDDIFDVLWQIAKKYEDNDKIIIFGLMNEBPHDIEIYAOTCOKVYTAIRKAGAT 171  
 Db 166 ----TDKAAFASFELKLTHTFASNPVLYVFDTNNEIHDMDQLVFLNDAALDAITAAAGT 241  
 OY 172 SOMILLPGTNFASVETVYVSGSAEALGRITNPDGSTDLLYFDVHKYITDINNNGSHAECT 231  
 Db 242 SOYIWEENSTWGATWMTVNTN--NLALRDEPK---LVYQMHOYLDDSGSTACVS 296  
 OY 232 DNV--DANDNPDADLROKKROAIISETGASMEPSCMTAFCAONKKAISENDVYIGFVWG 289  
 Db 297 TGVGLQRYVIGATNWLRONGKGLLGEFAGGANSVCQQAIEGLNLHLDENSDVTGALMWA 356  
 OY 290 AGSFDTSYILTLPLGKPGNYTDNKLINCEI 320  
 Db 357 GSPWMDYIYFSPSPSGIGYTYNSLKKYV 387  
 RESULT 8  
 cellulase (EC 3.2.1.4) precursor - imperfect fungus (Humicola insolens)  
 N:Alternate names: endo-1,4-beta-glucanase; endoglucanase  
 C:Species: Humicola insolens  
 C:Date: 11-Sep-1998 #sequence\_revision 11-Sep-1998 #text\_change 18-Jun-1999

C:Accession: S43920  
R:Dalboe, H.; Seidto-Hansen, H. P.  
Mol. Gen. Genet. 243, 253-260, 1994

A>Title: A novel method for efficient expression cloning of fungal enzyme genes.  
A:Reference number: S43919; MUID:94247364; PMID:8190078

A:Molecule type: mRNA  
A:Accession: S43920  
A:Residues: 1-388 <DAL>

A:Cross-references: EMBL:X76046; NID:9505194; PIDN:CMA53631.1; PID:9505195

C:Genetics:  
A:Gene: CMC3

C:Function:  
A:Description: hydrolysis of 1,4-beta-D-glucosidic linkages in beta-D-glucans such as  
A:Pathway: cellulose degradation  
C:Superfamily: Trichoderma cellulase III; fungal cellulose-binding domain homology  
C:Keywords: glycosidase, hydrolase, polysaccharide degradation  
F:1-16/Domains: signal sequence #status predicted <SIG>  
F:17-388/Product: cellulase #status predicted <MAT>  
F:21-52/Domains: fungal cellulose-binding domain homology <FCB>

Query Match 14.5%; Score 322; DB 1; Length 388;  
Best Local Similarity 26.6%; Pred. No. 7.9e-16;  
Matches 88; Conservative 54; Mismatches 151; Indels 38; Gaps 8;

```
OY      1 GKIKYLVAIPGIDFGCIDGSCP-----IDTSSVPLLSYKGGDAGOMKHFAEDD 51
DB      84 GKFKMGNGINQSAERG--KGEYPGLMGKHFFPTSSI-----QTHI--ND 125

OY      52 GLNVPRISATMOFVLNNTVDGKLDELNMGSYNKVVNAACLETGAYCMIDMHPARYNGII 111
DB     126 GNMRVRVASMERLAPNOLNAFAFDANYLRNLRETVNFIITGKGKTAMLPDPHNGRYETRI 185

OY     112 GGCGVSDDIFVDLWQAIRKYVEDNDKIIFGLMFEPHDIDIEIMAQTCOKVYTARIKAGAT 171
DB     186 ----TDKAAPAFSEFLTKLTHFASNPFLVEFDJNNEXHDDQQLVPDLNOALDAIRAACAT 241

OY     172 SQMLLPRTNFPAVSFTYYSTGSAAELGKITPNPDGSTDLLPEVNHLYIDDINNSGSHAECTT 231
DB     242 SOYLWEVNSWTGAWTMVTNN--NLAALRDPEKN--LVYOMHQYLDSDSGISTFACVS 296

OY     233 DNV--DAFNDFADMLROKRRKOAIISETGASMEPSCTAFCAONKAISESNPVYIGEVWG 289
DB     297 TGVGVQRIVIGATNMRLRGKVKGVLGEFPAGANSVCQAILEGTLHLQENSDVMWGMALMWA 356

OY     290 AGSEPDTYILTLTPLGKPGNTDNKLMNDCI 320
DB     357 GGPMMWDYIYSFEPPSGIGTYYNLSLKRVY 387
```

RESULT 9

JX0131  
cellulase (EC 3.2.1.4) - fungus (*Pestalotiopsis* sp.)  
N:Alternate names: endo-1,4-beta-glucanase; endo-cellulase  
C:Species: *Pestalotiopsis* sp.  
C>Date: 30-Jun-1992 #sequence\_revision 30-Jun-1992 #text\_change 24-Sep-1998  
C:Accession: JX0131; PS00333

R:yoshihi, N.; Taniguchi, H.; Sasaki, T.  
U:Biochem. 100, 388-392, 1990  
A>Title: Cloning and sequencing of the endo-cellulase cDNA from *Robillarda* sp. Y-20.  
A:Reference number: JX0131; MUID:91115784; PMID:2277031

A:Accession: JX0131  
A:Molecule type: mRNA  
A:Residues: 16-375 <YOS>  
A:Experimental source: strain Y-20  
A:Accession: PS00333

A:Molecule type: protein  
A:Residues: 1-18 <YOS1>  
A>Note: The authors translated the codon GGA for residue 353 as Trp

C:Function:  
A:Description: hydrolysis of 1,4-beta-D-glucosidic linkages in beta-D-glucans such as  
A:Pathway: cellulose degradation  
C:Superfamily: Pseudomonas solanacearum cellulase egl  
C:Keywords: glycoprotein; glycosidase; hydrolase; polysaccharide degradation



OY 59 SATQVFLNNTVDGKLDLDELNGSYNKVYNAACLETGAYCMIDMNFARNGGIIIGGGVSD 118  
 DB 153 PVGMQYLVNNNLGNDLSTSIKSYDOLVQGLSLGAYCIVIHMYARNGGIIIGGGPTN 212  
 OY 119 DIFVDLWVOIAKAYEDNDKIIIFGLMNEPHDIDIEIQAOTCKVYTAIRKAGATSMILLP 178  
 DB 213 AOFPSLSQSLASKYASOSRWFGIIMNEPHDINIMATVQEVYTAIRKAGATSOFLSLP 272  
 OY 179 GTFNASEVETVSTGSAALGKRTNPDGSTDLLYFVHXYLDINNSGSHAECTDNDV-AF 237  
 DB 273 GNDQSGAGAFISDGSAAALSVTPNDGSTNLIFDVHXYLDSDNSGTHAECTTNNDGAF 332  
 OY 238 NDFADMLRQNKROAIISFTGASMEPSCMTAFCAONKAISENSDVYIGFVWGAGSFDTSY 297  
 DB 333 SPLATWLRQNRQAIIETETGGVNOGCIQDWCOOIYVYLNQSDVYLYGVWGAGSFDSTY 392  
 OY 298 ILTLPGLKPGNYTNDKLNMECI 320  
 DB 393 VLTEPTSSGNSWTDTSLSVSSL 415

## RESULT 2

S68153  
 cellulase (EC 3.2.1.4) 3D precursor - Penicillium janthinellum  
 N:Alternate names: 3D endoglucanase 2; endo-1,4-beta-glucanase  
 C:Species: Penicillium janthinellum  
 C:Date: 11-Sep-1998 #sequence\_revision 11-Sep-1998 #text\_change 18-Jun-1999  
 R:Meritz, G.; Koch, A.; Henrissat, B.; Schulz, G.  
 Curr. Gene. 29, 490-495, 1996  
 A:Title: Endoglucanase II (EgII) of Penicillium janthinellum: cDNA sequence, heterologous  
 A:Reference number: S68153; MUID:96207475; PMID:8625430  
 A:Accession: S68153  
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A:Molecule type: mRNA  
 A:Residues: 1-410 <MR>  
 A:Cross-references: EMBL:X89564; NID:9984165; PIDN:CA61740.1; PID:9984166  
 A:Note: the nucleotide sequence was submitted to the EMBL Data Library, July 1995  
 C:Function:  
 A:Description: hydrolysis of 1,4-beta-D-glucosidic linkages in beta-D-glucans such as cel-  
 C:Superfamily: Trichoderma cellulase III; fungal cellulose-binding domain homology  
 C:Keywords: glycosidase; hydrolase; polysaccharide degradation  
 F:1-21/Domain: signal sequence #status predicted <SIG>  
 F:22-410/Product: cellulase #status predicted <MAT>  
 F:26-57/Domain: fungal cellulose-binding domain homology <FCB>

## Query Match

Best Local Similarity 40.08; Score 885.5; DB 1; Length 410;  
 Matches 163; Conservative 52; Mismatches 100; Indels 11; Gaps 3;

OY 1 GKIKYLVAIPIGIDGSCPTDTSVPLSYKGGDAGOMKHAEDDGLNVRISA 60  
 DB 92 GKVFAGVNTAGPFGVYVTS- - - - -TQDLSQVVDSEYDGNQSHVFNATFIFRPT 146  
 OY 61 TWQFVLNNTVDGKLDLDELNGSYNKVYNAACLETGAYCMIDMNFARNGGIIIGGGVSDI 120  
 DB 147 GMDIYVNNNLGSLDSNNFKYEQVGSGLSGAYCIVDINHVRMNGVYIGGGPDDQ 206  
 OY 121 FVDLWVOIAKAYEDNDKIIIFGLMNEPHDIDIEIQAOTCKVYTAIRKAGATSMILLPPT 180  
 DB 207 FISTLMTOLAHYKNSNVIFIGIMNEPHDINIMATVQEVYTAIRKAGATSMILLPPT 266  
 OY 181 NFASVETVSTGSAALGKRTNPDGSTDLLYFVHXYLDINNSGSHAECTDNDVDAFND 240  
 DB 267 DYTSAANFIENGSCAALLPTNPDGSTNLIFDVHXYLDSDNSGTHAECTTNNDGAFNL 326  
 OY 241 ADMLRQNKROAIISFTGASMEPSCMTAFCAONKAISENSDVYIGFVWGAGSFDTS- - -YI 298  
 DB 327 ATWLRQNKROAIISFTGASMEPSCMTAFCAONKAISENSDVYIGFVWGAGSFDTS- - -YI 298  
 OY 299 LTLPLGKPGNYTNDKLNMECI 320  
 DB 324 LTLPLGKPGNYTNDKLNMECI 324

DB 387 LLEVAV- - - - -NGVDQYLVQOCFVERKW 408

## RESULT 3

JC4115  
 endoglucanase precursor - Macrophomina phaseolina  
 C:Species: Macrophomina phaseolina  
 C:Date: 26-Jul-1995 #sequence\_revision 19-Oct-1995 #text\_change 24-Sep-1998  
 C:Accession: JC4115  
 R:Nang, H.; Jones, R.W.  
 Gene 158, 125-128, 1995  
 A:Title: Cloning, characterization and functional expression of an endoglucanase-enco  
 A:Reference number: JC4115; MUID:95309709; PMID:7789795  
 A:Accession: JC4115  
 A:Molecule type: mRNA  
 A:Residues: 1-338 <MR>  
 A:Cross-references: GB:U13914  
 C:Comment: This protein cleaves randomly within the cellulose chain, leading to a rap  
 C:Genetics:  
 A:Gene: eg12  
 C:Superfamily: Pseudomonas solanacearum cellulase eg1  
 C:Keywords: glycoprotein  
 F:1-19/Domain: signal sequence #status predicted <SIG>  
 F:20-338/Product: endoglucanase #status predicted <MAT>  
 F:222/Binding site: carbohydrate (Asn) (covalent) #status predicted

## Query Match

Best Local Similarity 33.18; Score 733; DB 2; Length 338;  
 Matches 143; Conservative 38; Mismatches 88; Indels 4; Gaps 4;

OY 2 KIKYLVAIPIGIDGSCPTDTSVPLSYKGGDAGOMKHAEDDGLNVRISA 61  
 DB 32 KVFAGVNTAGPFGVYVTS- - - - -TQDLSQVVDSEYDGNQSHVFNATFIFRPT 146  
 OY 62 WQFVLNNTVDGKLDLDELNGSYNKVYNAACLETGAYCMIDMNFARNGGIIIGGGVSDI 120  
 DB 91 WQYLVANNLGGDLDSANAGKYDNLVQGLSGAEICIIDINHVALLEPRDHROGPTNDQ 150  
 OY 121 FVDLWVOIAKAYEDNDKIIIFGLMNEPHDIDIEIQAOTCKVYTAIRKAGATSMILLP 179  
 DB 151 FVSLMKQATRTKNTKVAFAVNMNEPHDINIMATVQEVYTAIRKAGATSMILLP 210  
 OY 180 TNFASVETVSTGSAALGKRTNPDGSTDLLYFVHXYLDINNSGSHAECTDNDV-DAFN 238  
 DB 211 NDWLSAALFIDNGSAAALKKTNTPDGSTNLIFDVHXYLDSDNSGTHAECTTNNDGAF 270  
 OY 239 DFDMLRQNKROAIISFTGASMEPSCMTAFCAQ 271  
 DB 271 PLADMLRQNKRAIMTESGGGNTDSCERKFLVQ 303

## RESULT 4

JC1201  
 carboxymethylcellulase (EC 3.2.1.-) precursor - fungus (Cryptococcus flavus)  
 N:Alternate names: CMCase 1  
 C:Species: Cryptococcus flavus  
 C:Date: 30-Sep-1993 #sequence\_revision 30-Sep-1993 #text\_change 21-Jul-2000  
 C:Accession: JC1201  
 R:Chai, Z.; Mochizuki, D.; Matsuno, Y.; Nakamura, T.; Liu, Y.; Hatano, T.; Fukui, S.;  
 Biosci. Biotechnol. Biochem. 56, 1230-1235, 1992  
 A:Title: Cloning and molecular analysis of cDNA encoding a carboxymethylcellulase of  
 A:Reference number: JC1201; MUID:93005075; PMID:1368837  
 A:Accession: JC1201  
 A:Molecule type: DNA  
 A:Residues: 1-341 <CUT>  
 A:Cross-references: GB:U13967; NID:g217886; PIDN:BA03070.1; PID:g217887  
 C:Genetics:  
 A:Gene: CMCI  
 A:Initons: 41/3; 74/2; 140/3; 218/3; 252/2; 287/2  
 C:Superfamily: Pseudomonas solanacearum cellulase eg1  
 C:Keywords: glycosidase; hydrolase  
 F:1-16/Domain: signal sequence #status predicted <SIG>

GenCore version 5.1.4-p5-4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 9, 2003, 15:16:46 ; Search time 27.742 Seconds

(without alignments)  
1451.961 Million cell updates/sec

Title: US-10-028-245-2

Perfect score: 2214

Sequence: 1 GKIKYLGVAIPGIDFGCDID.....GLTGVLFVVALGYLVAF 419

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: PIR73:\*  
2: PIR1:\*  
3: PIR2:\*  
4: PIR4:\*

Pred. NO. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	942.5	42.6	418	1 S28372	cellulase (EC 3.2.
2	885.5	40.0	410	1 S68153	cellulase (EC 3.2.
3	733	33.1	338	2 JCA115	endoglucanase prec
4	654	29.5	341	2 JCA1201	carboxymethylcellu
5	387	17.5	356	2 G82523	endo-1,4-beta-gluc
6	386	17.4	426	2 A42649	cellulase (EC 3.2.
7	323	14.6	388	1 JCS461	cellulase (EC 3.2.
8	322	14.5	388	1 S43920	cellulase (EC 3.2.
9	217.5	9.8	375	2 JX0131	cellulase (EC 3.2.
10	193.5	8.7	814	1 C2CLEM	cellulase (EC 3.2.
11	186	8.4	1232	2 T31426	cellulase (EC 3.2.
12	175.5	7.9	473	2 S40507	endoglucanase - ru
13	165.5	7.4	482	2 JF0302	cellulase (EC 3.2.
14	158.5	7.2	547	2 S27500	cellulase (EC 3.2.
15	153	6.9	455	2 S16559	cellulase (EC 3.2.
16	153	6.9	722	2 H96136	endo-1,4-beta gluc
17	149.5	6.8	516	2 JF0194	mannan endo-1,4-be
18	147	6.6	417	2 B72216	endoglucanase - Th
19	147	6.6	309	2 B25156	cellulase (EC 3.2.
20	141	6.4	517	2 I40798	cellulase (EC 3.2.
21	138	6.2	656	2 AB1843	hypothetical prote
22	136	6.1	364	2 S12017	endoglucanase A -
23	135	6.0	475	1 C2CLCA	cellulase (EC 3.2.
24	132.5	6.0	377	1 F97001	endoglucanase fam1
25	129	5.8	584	2 S27500	xylanase - Prevore
26	128.5	5.8	1012	2 B97326	endoglucanase fam1
27	127.5	5.8	370	2 G97001	endoglucanase fam1
28	127.5	5.8	515	2 S20493	endoglucanase - Cl
29	124.5	5.6	1331	2 A48954	mannan endo-1,4-be

30	123.5	5.6	504	2 S54744	cellulase (EC 3.2.
31	123	5.6	557	2 JCS487	cellulase (EC 3.2.
32	123	5.6	566	2 A40589	cellulase (EC 3.2.
33	123	5.6	660	2 JMO067	chitinase (EC 3.2.
34	122	5.5	563	1 C2CLBM	cellulase (EC 3.2.
35	121.5	5.5	441	2 A44815	cellulase (EC 3.2.
36	120	5.4	363	2 I40234	carboxymethylcellu
37	117	5.3	806	2 E69424	hypothetical prote
38	116.5	5.3	499	2 A27198	cellulase (EC 3.2.
39	116.5	5.3	505	2 S39662	endoglucanase - Er
40	116.5	5.3	1063	2 D86731	hypothetical prote
41	116	5.2	4688	2 F82885	hypothetical prote
42	115.5	5.2	448	2 A27631	cellulase (EC 3.2.
43	115.5	5.2	1742	2 T17120	cellulase (EC 3.2.
44	115	5.2	584	2 J01229	cellulase (EC 3.2.
45	114.5	5.2	508	2 A26874	cellulase (EC 3.2.

## ALIGNMENTS

### RESULT 1

S28372 cellulase (EC 3.2.1.4) III precursor - fungus (Trichoderma reesei)

N:Alternate names: endo-1,4-beta-glucanase; endoglucanase III

C:Species: Trichoderma reesei

C>Date: 11-Sep-1998 #sequence, revision 11-Sep-1998 #extl\_change 18-Jun-1999

C:Accession: S28372; S78516; S78517; S02626

R:Salohelmo, M.; Lehtovaara, P.; Penttilae, M.; Teerit, T.T.; Stahlberg, J.; Johansson

Gene 63, 11-21, 1988

A:Title: EgIII, a new endoglucanase from Trichoderma reesei: the characterization of

A:Reference number: S28372; MUID:88255850; PMID:3384334

A:Accession: S28372

A:Molecule type: DNA

A:Residues: 1-418 <SAL>

A:Cross-references: EMBL:M19373; NID:g170548; PIDN:AAA34213.1; PID:g170549

A:Accession: S78516

A:Molecule type: mRNA

A:Residues: 1-418 <SAL2>

A:Cross-references: GB:M19373; NID:g170548; PIDN:AAA34213.1; PID:g170549

A:Accession: S78517

A:Molecule type: protein

A:Residues: 22-23, 'X', '25-46, 'XX', '49-58 <SAL3>

A:Stahlberg, J.; Johansson, G.; Petersson, G.

Eur. J. Biochem. 173, 179-183, 1988

A:Title: A binding-site-deficient, catalytically active, core protein of endoglucanase

A:Reference number: S02625; MUID:88185316; PMID:3356188

A:Accession: S02626

A:Molecule type: protein

A:Residues: 'XXXX', '87-88, 'V', '90, 'X', '92-93, 'Y', '95-99 <STA>

A:Experimental source: culture filtrate

C:Genetics:

A:Gene: egI3

A:Introns: 110/2

C:Function:

A:Description: hydrolysis of 1,4-beta-D-glucosidic linkages in beta-D-glucans such as

A:Pathway: cellulose degradation

C:Superfamily: Trichoderma cellulase III; fungal cellulose-binding domain homology

C:Keywords: blocked amino end; glycoprotein; glycosidase; hydrolase; polysaccharide d

F:1-21/Domain: signal sequence #status predicted <SIC>

F:22-418/Product: cellulase #status experimental <MAT1>

F:26-57/Domain: fungal cellulose-binding domain homology <FCB>

F:83-418/Product: cellulase core #status experimental <MAT2>

F:22/Modified site: pyroglutamate carboxylic acid (Glu) (in mature form) #status experi

F:124/Binding site: carboxylate (Asn) (covalent) #status predicted

Query Match 42.6%; Score 942.5; DB 1; Length 418;

Best local Similarity 51.7%; Pred. No. 8; 1e-61;

Matches 167; Conservative 61; Mismatches 90; Indels 5; Gaps 2;

DB 93 VRPAGVNIAGFDGCTTDCVTSKYVPLKNTGNNYVDDGIGOMOHFVNEGDMTIFRL.152

```

OY 163 TAIRKAGATSMILLP-----GTNEFASVETTYVSTGSAEALGKITNPDGSTDLLYFDVHK 216
Db 51 TTFHSGPGSTETILLPONTTASGLLEASTPYHSTGSPHT---TLSPAGSTT----- 99
OY 217 YLDINNGSHAECTTDVDAENDPADWLK-ONKROAISETGASME-----PSC--M 265
Db 100 -----ROGESTT-----FOSWPNKSDTTPAPPPTTSAFVELSTSHGSPSSSTPT 143
OY 266 TAPCAONKKAISENDVYIGFVGAGSFDTSYILTLPLGKPGNYTDNKLWNECILDQFT 325
Db 144 THFSASSTTLGRSEE-----STVYHSSPVATAT-TSPARSTISGLVEE----- 186
OY 326 LDERYRPTPTYSIST-----AAETAT--ATATSDGAPSTTKPIFREET---A 368
Db 187 -STVYHSSPGSTQTMHPESPSTTSGRGEESTTSHSTHTTISAPSTTSALVEEPTSYHS 245
OY 369 SPTFNAYTKPSPDSDSDDDKDSAASMSAGCLGTVL 406
Db 246 SPGSTATTH-FPDSSTTSGRSEESTASHSNODATGTTIV 282

```

Search completed: May 9, 2003, 15:17:20  
 job time : 36.5251 secs

FILING DATE: 10-OCT-1997  
APPLICATION NUMBER: 60/056,916  
FILING DATE: 06-DEC-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Hallie, Lisa A.  
REGISTRATION NUMBER: 38,347  
REFERENCE/DOCKET NUMBER: 09010/024002  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 858/677-1456  
TELEFAX: 858/677-1465  
INFORMATION FOR SEQ ID NO: 63:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 956 amino acids  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
FRAGMENT TYPE: internal  
SEQUENCE DESCRIPTION: SEQ ID NO: 63:  
US-10-121-032-63

Query Match 5.0%; Score 111.5; DB 9; Length 956;  
Best Local Similarity 20.9%; Pred. No. 0.76;  
Matches 75; Conservative 41; Mismatches 134; Indels 109; Gaps 15;

QY 78 NMGSYN-----KVVNACLETGAYCMIDMHNFRARYNGGIIQGGVSDIDFYD 123  
DB 97 NMDGGNGYIDSPQOEAKIRKVIDAANGIYIIDMHTH-----EAEIYDEAVD 147  
QY 124 LMVQIAKVEEDNDKIIIFGLMNEPHDDIEIWAQTCQKVYTAIRKAGATSMILLPTNFA 183  
DB 148 FETRMADLYGDTPNVMEIENEPIYOSWPYIKYAEQVLAGISKPPDN--LIIVGTSNV 205  
QY 184 SVETVYSTGSAEALGKITPNPDGSTDLLYFDVHKYLDIN--NSGSHAECTTDN----- 233  
DB 206 SQGVDAAS-----ADPISDTNVAY--TLHFYAFNPHDNLRNAQTALDNNVALFVY 255  
QY 234 -----VDAFNDFADMLRQNKROAIISETGASMEPSCMTAFCAONKAISENSDVIYGFVGW 288  
DB 256 EMGTILNTGGEPDKESTYTWMAFLKEKGIS-----HANWSLSDKAFPEGTGSVY--QAGQ 308  
QY 289 GAGSFPTSYILTLPLGKPGNVTDNKLMNECILDFTLDEKYPPTSTIST-----AAEE 343  
DB 309 GVSQ-----LISNKLITASGEIYKN-----IIQNMDETSTGPKTKTQCTSTECIRAAE 356  
QY 344 TATATATSDGD---AP-----STTKPIFREETASPTP 372  
DB 357 TAAQ-----GDEIIIAFGVNVNFQDKIQGAFNRSVLYXGANGNSTNPILLRGESATNP 410

RESULT 14  
US-10-093-037-63  
Sequence 63, Application US/10093037  
Publication No. US20030078397A1  
GENERAL INFORMATION:  
APPLICANT: Jay M. Short  
APPLICANT: Byline, Edward  
APPLICANT: Swanson, Ronald V.  
APPLICANT: Mathur, Eric J.  
APPLICANT: Lam, David E.  
TITLE OF INVENTION: ENZYMES HAVING GLYCOSIDASE ACTIVITY AND METHODS OF USE THEREOF  
FILE REFERENCE: 09010-024006  
CURRENT FILING DATE: US/10/093,037  
CURRENT FILING DATE: 2002-03-06  
PRIOR APPLICATION NUMBER: US 09/910,579  
PRIOR FILING DATE: 2001-07-20  
PRIOR APPLICATION NUMBER: US 09/134,078  
PRIOR FILING DATE: 1998-08-13  
PRIOR APPLICATION NUMBER: US 08/949,026  
PRIOR FILING DATE: 1997-10-10  
PRIOR APPLICATION NUMBER: US 60/056,916  
PRIOR FILING DATE: 1996-12-06  
NUMBER OF SEQ ID NOS: 72  
SOFTWARE: FASTSEQ for Windows Version 4.0

SEQ ID NO 63  
LENGTH: 956  
TYPE: PRT  
ORGANISM: Bankia gouldi  
US-10-093-037-63

Query Match 5.0%; Score 111.5; DB 9; Length 956;  
Best Local Similarity 20.9%; Pred. No. 0.76;  
Matches 75; Conservative 41; Mismatches 134; Indels 109; Gaps 15;

QY 78 NMGSYN-----KVVNACLETGAYCMIDMHNFRARYNGGIIQGGVSDIDFYD 123  
DB 97 NMDGGNGYIDSPQOEAKIRKVIDAANGIYIIDMHTH-----EAEIYDEAVD 147  
QY 124 LMVQIAKVEEDNDKIIIFGLMNEPHDDIEIWAQTCQKVYTAIRKAGATSMILLPTNFA 183  
DB 148 FETRMADLYGDTPNVMEIENEPIYOSWPYIKYAEQVLAGISKPPDN--LIIVGTSNV 205  
QY 184 SVETVYSTGSAEALGKITPNPDGSTDLLYFDVHKYLDIN--NSGSHAECTTDN----- 233  
DB 206 SQGVDAAS-----ADPISDTNVAY--TLHFYAFNPHDNLRNAQTALDNNVALFVY 255  
QY 234 -----VDAFNDFADMLRQNKROAIISETGASMEPSCMTAFCAONKAISENSDVIYGFVGW 288  
DB 256 EMGTILNTGGEPDKESTYTWMAFLKEKGIS-----HANWSLSDKAFPEGTGSVY--QAGQ 308  
QY 289 GAGSFPTSYILTLPLGKPGNVTDNKLMNECILDFTLDEKYPPTSTIST-----AAEE 343  
DB 309 GVSQ-----LISNKLITASGEIYKN-----IIQNMDETSTGPKTKTQCTSTECIRAAE 356  
QY 344 TATATATSDGD---AP-----STTKPIFREETASPTP 372  
DB 357 TAAQ-----GDEIIIAFGVNVNFQDKIQGAFNRSVLYXGANGNSTNPILLRGESATNP 410

RESULT 15  
US-10-025-380-1065  
Sequence 1065, Application US/10025380  
Publication No. US20020182191A1  
GENERAL INFORMATION:  
APPLICANT: Xu, Jiangchun  
APPLICANT: Lodes, Michael J.  
APPLICANT: Secrist, Heather  
APPLICANT: Benson, Darin R.  
APPLICANT: Meagher, Madeleine Joy  
APPLICANT: Stolk, John A.  
APPLICANT: Wang, Tongtong  
APPLICANT: Jiang, Yugu  
APPLICANT: Smith, Carole L.  
APPLICANT: King, Gordon E.  
APPLICANT: Wang, Aijun  
APPLICANT: Clapper, Jonathan D.  
APPLICANT: Skelky, Yasir A. W.  
APPLICANT: Fanger, Gary R.  
APPLICANT: Vedyck Thomas S.  
APPLICANT: Carter, Darick  
TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY AND DIAGNOSIS  
FILE REFERENCE: 210121.471C14  
CURRENT APPLICATION NUMBER: US/10/025,380  
CURRENT FILING DATE: 2001-12-19  
NUMBER OF SEQ ID NOS: 1129  
SOFTWARE: FASTSEQ for Windows Version 4.0  
SEQ ID NO 1065  
LENGTH: 957  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-10-025-380-1065

Query Match 5.0%; Score 111; DB 9; Length 957;  
Best Local Similarity 24.1%; Pred. No. 0.83;  
Matches 67; Conservative 31; Mismatches 100; Indels 80; Gaps 14;

;; TITLE OF INVENTION: Modified Target Enzymes Having Catalytic  
;; TITLE OF INVENTION: Trials  
;; FILE REFERENCE: and Uses Thereof  
;; CURRENT APPLICATION NUMBER: US/10/068,374  
;; CURRENT FILING DATE: 2002-02-06  
;; PRIOR APPLICATION NUMBER: US 60/268,347  
;; PRIOR FILING DATE: 2001-02-12  
;; NUMBER OF SEQ ID NOS: 5  
;; SOFTWARE: FastSeq for Windows Version 4.0  
;; SEQ ID NO 5  
;; LENGTH: 308  
;; TYPE: PRT  
;; ORGANISM: Artificial Sequence  
;; FEATURE:  
;; OTHER INFORMATION: protein encoded by plasmid pCORE3  
US-10-068-374-5

Query Match 5.3%; Score 118; DB 9; Length 308;  
Best Local Similarity 19.2%; Pred. No. 0.044;  
Matches 61; Conservative 58; Mismatches 123; Indels 76; Gaps 13;

QY 44 MKHAEEDGLNFR---TSATMOFLNNTVDGKDELNMGSYKVNACLEFGAYCMTDM 100  
DB 49 MKMLRDMGIVFRAAMITSSGGYIDPSYKKEKE-----TVEAIDGITYIIDM 100  
QY 101 HNFARYNGIIGGCGVSDIDFVDMVQIAKYEDNDKIIFGLMNEPHDLIEIWAQT--- 157  
DB 101 HILSDNDPIYKEEA-----KDFEDEMSELYGDPVNIYELANPNQSDV-TWNOQIKP 153  
QY 158 -CQKVTYIRKAGATSCMILLPGTNFASVETVYSGSAEALGKTIKINPDGSDLYFDVHK 216  
DB 154 YAEVETPIYIRNDPN--TVIVGTGTSQDVHNA---ADNOLDPN---WYAFHF 201  
QY 217 YLDINNSGSHAECTDNDVAFNDFADMLRQNKRAIISGTSMPSCMTAFCAQNKAIIS 276  
DB 202 Y-----ACFHQNLRDY-----DYALDGAALTYSENGTS-----NATGDGVFL 242  
QY 277 ENSDVIYGFV-----GNGAGSEFDSYILTLPLGKPGVNTDKMLNECILDQFTLDE 328  
DB 243 DEAOVWIDFMDERNLSWAMNSLTHHDESSAALMPCANPTGGTEAEL----- 289  
QY 329 KYRPTPSISTAAEETAT 346  
DB 290 --SPSGTFVRKIREASAS 305

RESULT 12  
US-09-801-368-82  
;; Sequence 82, Application US/09801368  
;; Patent No. US20020128250A1  
;; GENERAL INFORMATION:  
;; APPLICANT: Busby, Robert  
;; APPLICANT: Call, Brian  
;; APPLICANT: Hecht, Peter  
;; APPLICANT: Holtzman, Doug  
;; APPLICANT: Madden, Kevin  
;; APPLICANT: Maxon, Mary  
;; APPLICANT: Milne, Todd  
;; APPLICANT: No. US20020128250A1man, Thea  
;; APPLICANT: Royer, John  
;; APPLICANT: Salama, Sofie  
;; APPLICANT: Sherman, Amir  
;; APPLICANT: Silva, Jeff  
;; APPLICANT: Summers, Eric  
;; TITLE OF INVENTION: Methods for Improving Secondary Metabolite Production in Fungi  
;; FILE REFERENCE: 109272.147  
;; CURRENT APPLICATION NUMBER: US/09/801,368  
;; FILING DATE: 2001-03-07  
;; PRIOR APPLICATION NUMBER: US 09/487,558  
;; PRIOR FILING DATE: 2000-01-19  
;; PRIOR APPLICATION NUMBER: US 60/160,587  
;; PRIOR FILING DATE: 1999-10-20

;; NUMBER OF SEQ ID NOS: 440  
;; SOFTWARE: PatentIn version 3.0  
;; SEQ ID NO 82  
;; LENGTH: 562  
;; TYPE: PRT  
;; ORGANISM: Saccharomyces cerevisiae  
US-09-801-368-82

Query Match 5.1%; Score 112; DB 10; Length 562;  
Best Local Similarity 20.1%; Pred. No. 0.33;  
Matches 95; Conservative 63; Mismatches 141; Indels 174; Gaps 24;

QY 24 PTD-----TSSVPLLSYKGGDAGQMKHFA---EDGLNVRISATWQFLNNTVDGKIDE 76  
DB 17 PTDAFDRSANTNINIAVWQNSAGTQESLATYCESSDADIFILSFLNPTLG----- 68  
QY 77 LNWGSYKVNVAACLETGAYCAIDMHNFRYNGIIGGCGVSDIDFVDMVQIAKYEDND 136  
DB 69 LNF-----ANAGSDT-----FSDGLHCTQIAEDI----- 93  
QY 137 KIIFGLMNEPHDLIEIWAQTCQ---KVTAFIRKAGATSCMILLPGTNFASVETVYST- 191  
DB 94 -----ETCQSLGKRVLLSL--GASGSYLF--SDSQAEFFAOTL 129  
QY 192 --GSAEALGKTIKTPDGSITDLYFDVHKYLDINNSGSHAECTDNDVAFNDFADMLR--- 245  
DB 130 WDFEGEGTAGSERPDSAYVDGFD---FDIEN-----NNEVGSALATKIRTLFA 176  
QY 246 QNKRQALISFEGASBPSCMTAFCAQNKAISENSDVIYIGFGW-----GASSFDI- 295  
DB 177 EGTQOYLYSAA---PCQPYPDASVGDLL-ENMDIDPAFIQFYNNYCSVSGQFNMDTWL 230  
QY 296 SYILTLFP-----LGKPGN-----TYDNKLMNECIUD-----Q 323  
DB 231 TYAQVSPNKNKIKLFLGIPGASAAAGSCYISDTILSTIADIASSSFGIALMDASQA 290  
QY 324 FTLD--EKYRPTPSISTAAEETATAT-ATSDGAPSTTKPIF----- 363  
DB 291 FSNELNGEPYVEILKLLTSSAQATATTVATRSKTSASSTASSTASKTKTQSTTS 350  
QY 364 ---RETFASPTPAVNTKPS---PDRSDSSDDDKDSAAASAGLICITVLT 408  
DB 351 TQSKSKVTLSPASSAATKTSITQTTKTLTSSTKTKSSLGTTTSTLNSVAIT 403

RESULT 13  
US-10-121-032-63  
;; Sequence 63, Application US/10121032  
;; Patent No. US20020155550A1  
;; GENERAL INFORMATION:  
;; APPLICANT: Bylina, Edward J.  
;; TITLE OF INVENTION: GLYCOSIDASE ENZYMES  
;; NUMBER OF SEQUENCES: 72  
;; CORRESPONDENCE ADDRESS:  
;; ADDRESSEE: Gray Cary Ware & Freidenrich LLP  
;; STREET: 4365 Executive Drive, Suite 1600  
;; CITY: San Diego  
;; STATE: CA  
;; COUNTRY: USA  
;; ZIP: 92121  
;; COMPUTER READABLE FORM:  
;; MEDIUM TYPE: Diskette  
;; COMPUTER: IBM Compatible  
;; OPERATING SYSTEM: Windows95  
;; SOFTWARE: FastSeq for Windows Version 2.0  
;; CURRENT APPLICATION DATA:  
;; APPLICATION NUMBER: US/10/121,032  
;; FILING DATE: 09-Apr-2002  
;; CLASSIFICATION: <Unknown>  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: US/09/134,078  
;; FILING DATE: 13-AUG-1998  
;; APPLICATION NUMBER: 08/949,026

Best Local Similarity 19.8%; Pred. No. 0.0046;  
Matches 67; Conservative 58; Mismatches 132; Indels 81; Gaps 14;

```

OY 44 MKHFAEDDGLNFR---ISATWQFVLNNTVDGKLDELNMGSYKYNVACLETGAYCMIDM 100
    || :||:||||| :||: :||: :||: :||: :||: :||: :||: :||:
DB 75 MKWLKRDWGTIVFRAMNYTSSGGYIDDPYKEKYKE-----AVEALDILGIYIIDW 126

OY 101 HNFARYNGIIGGVSDDIFVDLWQIAKYEDNDKIIIFGLNNEPHDDIEIWAQT--- 157
    | :||: :||: :||: :||: :||: :||: :||: :||: :||:
DB 127 HILSDNDPNLYKEA-----KDFEEMSELGYDIPNVIEIANERNGSDV-TWNOQIKP 179

OY 158 -COKVYTAIRKAGATSONILLPCTNFASVETVYSGSAEALGKITNPDGSTDLLYFDVHK 216
    :||: :||: :||: :||: :||: :||: :||: :||: :||:
DB 180 YAEVPIIRNNDPN--IIYVGTGWSQDVHHA-----ADNOLADPN-----VMYAFHF 227

OY 217 YLDINNSGSHAECTDNVDAFNDADMLRQNKROAIISSETGASMEPSCHMTAFCAQNKAIS 276
    | :||: :||: :||: :||: :||: :||: :||: :||:
DB 228 Y-----AGTHGQNLROV-----DYALDQGAIFVSEWGT-----AATGDCGVFL 268

OY 277 ENSDYVIGFY-----GWSGSEPTSYLLITPLGKPGNYDNKLMNECILDQFTLDE 328
    :||: :||: :||: :||: :||: :||: :||: :||:
DB 269 DEAOVWIDFDERNLNLSWAMNSLTHKDESSAALMPGANPTGCGWTEAEI----- 315

OY 329 KYRPTPTSISTAETAT-----ATATSDGADASTTKP 361
    | :||: :||: :||: :||: :||: :||: :||:
DB 316 --SPSGTFVREKIRRESASIPSPDPTPPDPGPEPTPP 351

```

## RESULT 9

US-10-068-374-2  
; Sequence 2, Application US/10068374  
; Publication No. US20030044903A1  
; GENERAL INFORMATION:

```

; APPLICANT: Shaw, Andrew
; TITLE OF INVENTION: Modified Target Enzymes Having Catalytic
; TITLE OF INVENTION: and Uses Thereof
; FILE REFERENCE: GC652
; CURRENT APPLICATION NUMBER: US/10/068,374
; CURRENT FILING DATE: 2002-02-06
; PRIOR APPLICATION NUMBER: US 60/268,347
; PRIOR FILING DATE: 2001-02-12
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 467
; TYPE: PRT
; ORGANISM: Bacillus sp.
US-10-068-374-2

```

Query Match  
Best Local Similarity 19.8%; Pred. No. 0.0063;  
Matches 69; Conservative 59; Mismatches 133; Indels 88; Gaps 15;

```

OY 44 MKHFAEDDGLNFR---ISATWQFVLNNTVDGKLDELNMGSYKYNVACLETGAYCMIDM 100
    || :||:||||| :||: :||: :||: :||: :||: :||: :||:
DB 75 MKWLKRDWGTIVFRAMNYTSSGGYIDDPYKEKYKE-----TVEAALDILGIYIIDW 126

OY 101 HNFARYNGIIGGVSDDIFVDLWQIAKYEDNDKIIIFGLNNEPHDDIEIWAQT--- 157
    | :||: :||: :||: :||: :||: :||: :||: :||:
DB 127 HILSDNDPNLYKEA-----KDFEEMSELGYDIPNVIEIANERNGSDV-TWNOQIKP 179

OY 158 -COKVYTAIRKAGATSONILLPCTNFASVETVYSGSAEALGKITNPDGSTDLLYFDVHK 216
    :||: :||: :||: :||: :||: :||: :||: :||:
DB 180 YAEVPIIRNNDPN--IIYVGTGWSQDVHHA-----ADNOLADPN-----VMYAFHF 227

OY 217 YLDINNSGSHAECTDNVDAFNDADMLRQNKROAIISSETGASMEPSCHMTAFCAQNKAIS 276
    | :||: :||: :||: :||: :||: :||: :||:
DB 228 Y-----AGTHGQNLROV-----DYALDQGAIFVSEWGT-----AATGDCGVFL 268

OY 277 ENSDYVIGFY-----GWSGSEPTSYLLITPLGKPGNYDNKLMNECILDQFTLDE 328
    :||: :||: :||: :||: :||: :||: :||:
DB 269 DEAOVWIDFDERNLNLSWAMNSLTHKDESSAALMPGANPTGCGWTEAEI----- 315

```

```

OY 329 KYRPTPTSISTAETAT-----ATATSDGADASTTKPIPRETASPTP 372
    | :||: :||: :||: :||: :||: :||: :||:
DB 316 --SPSGTFVREKIRRESASIPSPDPTPPDPGPEPTPP-----DPTP 355

```

## RESULT 10

US-09-863-547B-1  
; Sequence 1, Application US/09863547B  
; Patent No. US20020128166A1  
; GENERAL INFORMATION:

```

; APPLICANT: Henkel KgaA
; APPLICANT: Hermannus, Lentus B. M.
; APPLICANT: Van Beckhoven, Rudolf F. W. C.
; APPLICANT: Maurer, Karl-Helinz
; APPLICANT: Kottwitz, Beatrix
; APPLICANT: Weiss, Albrecht
; APPLICANT: Van Solingen, Pieter
; TITLE OF INVENTION: Detergents Comprising Cellulases
; FILE REFERENCE: H 1920 A
; CURRENT APPLICATION NUMBER: US/09/863,547B
; PRIOR FILING DATE: 2002-02-13
; PRIOR APPLICATION NUMBER: US 08/945,574
; PRIOR FILING DATE: 1998-02-27
; PRIOR APPLICATION NUMBER: PCT/EP96/01755
; PRIOR FILING DATE: 1996-04-26
; PRIOR APPLICATION NUMBER: US 08/614,115
; PRIOR FILING DATE: 1996-03-12
; PRIOR APPLICATION NUMBER: EP 95201115.3
; PRIOR FILING DATE: 1995-04-28
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
; LENGTH: 467
; TYPE: PRT
; ORGANISM: Bacillus sp. CBS 670.93
US-09-863-547B-1

```

Query Match  
Best Local Similarity 19.8%; Pred. No. 0.0063;  
Matches 69; Conservative 59; Mismatches 133; Indels 88; Gaps 15;

```

OY 44 MKHFAEDDGLNFR---ISATWQFVLNNTVDGKLDELNMGSYKYNVACLETGAYCMIDM 100
    || :||:||||| :||: :||: :||: :||: :||: :||: :||:
DB 75 MKWLKRDWGTIVFRAMNYTSSGGYIDDPYKEKYKE-----TVEAALDILGIYIIDW 126

OY 101 HNFARYNGIIGGVSDDIFVDLWQIAKYEDNDKIIIFGLNNEPHDDIEIWAQT--- 157
    | :||: :||: :||: :||: :||: :||: :||:
DB 127 HILSDNDPNLYKEA-----KDFEEMSELGYDIPNVIEIANERNGSDV-TWNOQIKP 179

OY 158 -COKVYTAIRKAGATSONILLPCTNFASVETVYSGSAEALGKITNPDGSTDLLYFDVHK 216
    :||: :||: :||: :||: :||: :||: :||:
DB 180 YAEVPIIRNNDPN--IIYVGTGWSQDVHHA-----ADNOLADPN-----VMYAFHF 227

OY 217 YLDINNSGSHAECTDNVDAFNDADMLRQNKROAIISSETGASMEPSCHMTAFCAQNKAIS 276
    | :||: :||: :||: :||: :||: :||: :||:
DB 228 Y-----AGTHGQNLROV-----DYALDQGAIFVSEWGT-----AATGDCGVFL 268

OY 277 ENSDYVIGFY-----GWSGSEPTSYLLITPLGKPGNYDNKLMNECILDQFTLDE 328
    :||: :||: :||: :||: :||: :||: :||:
DB 269 DEAOVWIDFDERNLNLSWAMNSLTHKDESSAALMPGANPTGCGWTEAEI----- 315

OY 329 KYRPTPTSISTAETAT-----ATATSDGADASTTKPIPRETASPTP 372
    | :||: :||: :||: :||: :||: :||: :||:
DB 316 --SPSGTFVREKIRRESASIPSPDPTPPDPGPEPTPP-----DPTP 355

```

## RESULT 11

US-10-068-374-5  
; Sequence 5, Application US/10068374  
; Publication No. US20030044903A1  
; GENERAL INFORMATION:

```

; APPLICANT: Shaw, Andrew

```

Db 100 -LAMDPEHKEFLAKQADRYKDPETLFEELNEPBNLPEKMWELLEEALVIR 158  
QY 167 KACATSOHMLPGTNFASVETVYSTGSAEALGKTITPDGSDLLFVHYKLOINNSGH 226  
Db 159 SIDKHHITII--GT-----AEMGGSIALEKISVPEKERNST-VTHIHYNPEEFTHOG 207  
QY 227 AECTTDN-----VDAFNDFADMLRONKROAI;SETGASNEPSCMTAF 269  
Db 208 AEWEGSEKMLGRKMGSPDDOKHLEEFNLEEMSKKNRPYIGERG----- 255  
QY 270 AOKKAISENDVYIGFV-----GW-----GAGSEDT 295  
Db 256 AYKRADESIRIKWTSFYVREMERKRWMSWAYEFCGFDY 297

RESULT 7  
US-09-903-185-2  
Sequence 2, Application US/09903185  
Publication No. US20030054534A1  
GENERAL INFORMATION:  
APPLICANT: Outtrup, Helle  
Olsen, Arne  
Bisg rd-Frantzen, Henrik  
Sch lein, Martin  
J rgensen, Per  
Bjoernvad, Mads  
TITLE OF INVENTION: DNA Constructs and Methods of Producing  
Cellulytic Enzymes- 1  
NUMBER OF SEQUENCES: 14  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: NO. US20030054534A10 No. US20030054534A1disk of No. US20030054534A1  
STREET: 405 Lexington Avenue, Suite 6400  
CITY: New York  
STATE: New York  
COUNTRY: U.S.A.  
ZIP: 10174-6401  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25 (EPO)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/903,185  
FILING DATE: 11-Jul-2001  
CLASSIFICATION: <Unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/870,180  
FILING DATE: 6-June-1997  
ATTORNEY/AGENT INFORMATION:  
NAME: Rozek, Carol E.  
REGISTRATION NUMBER: 36,993  
REFERENCE/DOCKET NUMBER: 3794,434-US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 212 867 0123  
TELEFAX: 212 867 0298  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 400 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
SEQUENCE DESCRIPTION: SEQ ID NO: 2:  
US-09-903-185-2

Query Match 6.2%; Score 136.5; DB 9; Length 400;  
Best Local Similarity 19.9%; Pred. No. 0.0017;  
Matches 70; Conservative 60; Mismatches 14; Indels 81; Gaps 14;  
QY 44 MKHFAEDDGINVFR---ISATMGVLNVTGDKDELNMGSYNNVYVACLETGAYCHIDM 100  
Db 75 MKLURDWDGINVFRAMMTSSGGTIDDPVSKYKVE-----AVERAIDLDIVYIIDW 126

QY 101 HNFARNGGIIGOGVSDIDFVLMVOIARYEDNDKIIIFGLNPEPHLDIEIWAOT--- 157  
Db 127 HILSDNPNYIKKEA-----KOFDEMSSELYGYPRVYIEIANEPNGSDV-TWGNOIKP 179  
QY 158 -CQKVTAIRKAGATSCMILLPGTNFASVETVYSTGSAEALGKTITPDGSDLLFVHYK 216  
Db 180 YAEVPIIIRNPNPN--IIIVGTGTSODVHNA-----ADNQLADPN-----VMAFHF 227  
QY 217 YLDINNGSHAECTDVDAFNDFADMLRONKRAIISFGASNEPSCMTAFCONKAIS 276  
Db 228 Y-----AGTHGQNLRDVY-----DYALDGAALFVSEMGTS-----AATGDCGVFL 268  
QY 277 ENSDVIYGFV-----GNGAGSEDTSYIILTPPLKPGNYTDNKLINCEILDQFTLDE 328  
Db 269 DEAGVWIDPFMDERNLISWANNSLTHKDESSAALMPGAMPPTGGWTEAEL----- 315  
QY 329 KYRPTPSISTAAEETAT-----ATATSDGDAPESTYPIRFEETASPTPNAV 375  
Db 316 --SPSGTFVREKIREASAIIPSDPTPPSDPDGEPDPTPSDGPDPADDPNOI 365

RESULT 8  
US-09-903-185-13  
Sequence 13, Application US/09903185  
Publication No. US20030054534A1  
GENERAL INFORMATION:  
APPLICANT: Outtrup, Helle  
Olsen, Arne  
Bisg rd-Frantzen, Henrik  
Sch lein, Martin  
J rgensen, Per  
Bjoernvad, Mads  
TITLE OF INVENTION: DNA Constructs and Methods of Producing  
Cellulytic Enzymes  
NUMBER OF SEQUENCES: 14  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: NO. US20030054534A10 No. US20030054534A1disk of No. US20030054534A1  
STREET: 405 Lexington Avenue, Suite 6400  
CITY: New York  
STATE: New York  
COUNTRY: U.S.A.  
ZIP: 10174-6401  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25 (EPO)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/903,185  
FILING DATE: 11-Jul-2001  
CLASSIFICATION: <Unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/870,180  
FILING DATE: 6-June-1997  
ATTORNEY/AGENT INFORMATION:  
NAME: Rozek, Carol E.  
REGISTRATION NUMBER: 36,993  
REFERENCE/DOCKET NUMBER: 3794,434-US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 212 867 0123  
TELEFAX: 212 867 0298  
INFORMATION FOR SEQ ID NO: 13:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 462 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
SEQUENCE DESCRIPTION: SEQ ID NO: 13:  
US-09-903-185-13

Query Match 6.0%; Score 132.5; DB 9; Length 462;





GenCore version 5.1.4-p5.4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 9, 2003, 15:09:30 ; Search time 32.5251 Seconds

(without alignments)  
1185,508 Million cell updates/sec

Title: US-10-028-245-2

Perfect score: 2214

Sequence: 1 GKIRYLGVAIPGIDFGDID.....GIGTGVLTVAALGYMLVAF 419

Scoring table: BLOSUM62

Searched: 349150 seqs, 92025710 residues

Total number of hits satisfying chosen parameters: 349150

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published\_Applications\_AA:\*

1: /cgn2\_6/ptodata/2/pubpaa/US08\_NEW\_PUB.pep:\*

2: /cgn2\_6/ptodata/2/pubpaa/PCIT\_NEW\_PUB.pep:\*

3: /cgn2\_6/ptodata/2/pubpaa/US06\_NEW\_PUB.pep:\*

4: /cgn2\_6/ptodata/2/pubpaa/US06\_PUBCOMB.pep:\*

5: /cgn2\_6/ptodata/2/pubpaa/US07\_NEW\_PUB.pep:\*

6: /cgn2\_6/ptodata/2/pubpaa/US07\_PUBCOMB.pep:\*

7: /cgn2\_6/ptodata/2/pubpaa/PCITUS\_PUBCOMB.pep:\*

8: /cgn2\_6/ptodata/2/pubpaa/US08\_PUBCOMB.pep:\*

9: /cgn2\_6/ptodata/2/pubpaa/US09\_NEW\_PUB.pep:\*

10: /cgn2\_6/ptodata/2/pubpaa/US09\_PUBCOMB.pep:\*

11: /cgn2\_6/ptodata/2/pubpaa/US10\_NEW\_PUB.pep:\*

12: /cgn2\_6/ptodata/2/pubpaa/US10\_PUBCOMB.pep:\*

13: /cgn2\_6/ptodata/2/pubpaa/US60\_NEW\_PUB.pep:\*

14: /cgn2\_6/ptodata/2/pubpaa/US60\_PUBCOMB.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	942.5	42.6	327	US-09-916-494A-16	Sequence 16, Appl
2	383.5	17.3	333	US-09-927-827-51	Sequence 51, Appl
3	376.5	17.0	335	US-09-927-827-52	Sequence 52, Appl
4	146.5	6.6	360	US-10-121-032-27	Sequence 27, Appl
5	146.5	6.6	360	US-10-093-037-27	Sequence 27, Appl
6	145	6.5	317	US-09-880-729-2	Sequence 2, Appl
7	136.5	6.2	400	US-09-903-185-2	Sequence 2, Appl
8	132.5	6.0	462	US-09-903-185-13	Sequence 13, Appl
9	131	5.9	467	US-10-068-374-2	Sequence 2, Appl
10	131	5.9	467	US-09-863-547B-1	Sequence 1, Appl
11	118	5.3	308	US-10-068-374-5	Sequence 5, Appl
12	112	5.1	562	US-09-801-368-82	Sequence 82, Appl
13	111.5	5.0	956	US-10-121-032-63	Sequence 63, Appl
14	111.5	5.0	956	US-10-093-037-63	Sequence 63, Appl
15	111	5.0	957	US-10-025-380-1065	Sequence 1065, Ap
16	111	5.0	957	US-09-922-217-1065	Sequence 1065, Ap
17	111	5.0	957	US-09-833-263-1065	Sequence 1065, Ap
18	110.5	5.0	1367	US-09-801-368-108	Sequence 108, App
19	109.5	4.9	2344	US-09-815-242-12713	Sequence 12713, A

20	108.5	4.9	502	US-09-815-242-5904	Sequence 5904, Ap
21	108.5	4.9	553	US-09-888-224-2	Sequence 2, Appl
22	108.5	4.9	560	US-09-815-242-13057	Sequence 13057, A
23	106	4.8	562	US-09-981-900B-5	Sequence 5, Appl
24	105.5	4.8	574	US-09-863-547B-2	Sequence 2, Appl
25	105.5	4.8	1621	US-10-185-990-10	Sequence 10, Appl
26	104.5	4.7	762	US-10-163-547-1	Sequence 1, Appl
27	100	4.5	333	US-09-927-827-3	Sequence 3, Appl
28	99.5	4.5	348	US-09-797-464A-2	Sequence 2, Appl
29	97.5	4.4	348	US-09-797-464A-7	Sequence 7, Appl
30	97	4.4	345	US-09-953-280-37	Sequence 37, Appl
31	97	4.4	550	US-10-123-692-22	Sequence 22, Appl
32	96.5	4.4	464	US-09-815-242-10647	Sequence 10647, A
33	96.5	4.4	467	US-09-815-242-1997	Sequence 4997, Ap
34	96.5	4.4	5179	US-10-025-380-1068	Sequence 1068, Ap
35	96.5	4.4	5179	US-09-922-217-1068	Sequence 1068, Ap
36	96.5	4.4	5179	US-09-833-263-1068	Sequence 1068, Ap
37	96	4.3	329	US-09-927-827-16	Sequence 16, Appl
38	96	4.3	503	US-09-801-368-424	Sequence 424, App
39	96	4.3	990	US-10-101-464A-814	Sequence 814, App
40	95.5	4.3	773	US-10-067-385-8	Sequence 8, Appl
41	95.5	4.3	1031	US-09-815-242-10932	Sequence 10932, A
42	95.5	4.3	1290	US-09-881-752A-138	Sequence 138, App
43	94.5	4.3	341	US-09-953-280-36	Sequence 36, Appl
44	94.5	4.3	352	US-09-978-295A-612	Sequence 612, Appl
45	94.5	4.3	352	US-09-978-697-612	Sequence 612, App

## ALIGNMENTS

RESULT 1

US-09-916-494A-16

Sequence 16, Application US/09916494A

Patent No. US2002016477A1

GENERAL INFORMATION:

APPLICANT: Fowler, Timothy

APPLICANT: Clarkson, Kathleen A.

APPLICANT: Ward, Michael

APPLICANT: Collier, Katherine D.

TITLE OF INVENTION: Method and Compositions for Treating

TITLE OF INVENTION: Cellulose Containing Fabrics Using Truncated Cellulase

FILE REFERENCE: GC226-C4

CURRENT APPLICATION NUMBER: US/09/916,494A

CURRENT FILING DATE: 2000-06-14

PRIOR APPLICATION NUMBER: US 08/382,452

PRIOR FILING DATE: 1995-02-01

PRIOR APPLICATION NUMBER: US 08/169,948

PRIOR FILING DATE: 1993-12-17

NUMBER OF SEQ ID NOS: 43

SOFTWARE: FastSeq for Windows Version 4.0

SEQ ID NO 16

LENGTH: 327

TYPE: PRT

ORGANISM: Trichoderma longibrachiatum

US-09-916-494A-16

Query Match

Best Local Similarity 51.7%; Pred. No. 9.7e-72;

Matches 167; Conservative 61; Mismatches 90; Indels 5; Gaps 2;

3 IYLVGAIPGIDFGDIDSCPTDSSVPLLSYKGS-----DAGQMKHFAEDGILNVPRI 58

2 VFAGVNIAGFGFCCTDCTVTSKYVPLKNFTSSNNYPIDSGMQHFNVDGMITFEL 61

59 SATWQVLTNTDQKLDLELNMGSYKYNVACLETGAYCVIDMHNFRARYNGITGCGVSD 118

62 PIGWQVLTNNIGGNDLSISIKSLQVGGCLSLGAYCIVDHNFRARYNGITGCGVSD 121

119 DIFVDLWQVIAKYEDNDKIIIFGLNNEPHDIDIEIWAQTCQKRVYAIRRAGATSQMILP 178

**THIS PAGE BLANK (USPTO)**



QY 168 AGA--TSOMILLPGTNPASVETVSTGSAEALGKITNPDGSTDLLYEDVHKYLDIN--N 222  
 Db 208 TGGNNARHLMIP-----TYAASVNDGGINNFKYPNNGD-DKVIYSLHSYSPNFALN 258  
 QY 223 SGSHAETTDNVDAFND--FADWLROKROAIISSETGASMEPSCMTAFCAONKAISESD 280  
 Db 259 NGPGA-----ISNFYGENEIDWVMNTINSSFSISKG---IPVILIGEFAVAMNR--DNED 305  
 QY 281 VYIGFVGAGSFPDTSYLLTLPKPKPGNTYDN 313  
 Db 306 -----DREMOEYIKKATALGIPCVIMDN 330

RESULT 11  
 US-09-687-147-12  
 ; Sequence 12, Application US/09687147  
 ; Patent No. 6268198  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Li, Xin-Liang  
 ; APPLICANT: Ljungdahl, Lars G.  
 ; APPLICANT: Chen, Huizhong  
 ; TITLE OF INVENTION: Cellulases and Coding Sequences  
 ; FILE REFERENCE: 42-96a  
 ; CURRENT APPLICATION NUMBER: US/09/687,147  
 ; CURRENT FILING DATE: 2000-10-12  
 ; PRIOR APPLICATION NUMBER: US 60/027, 883  
 ; PRIOR FILING DATE: 1996-10-04  
 ; PRIOR APPLICATION NUMBER: PCT US97/18008  
 ; PRIOR FILING DATE: 1997-10-03  
 ; PRIOR APPLICATION NUMBER: 09/286,691  
 ; PRIOR FILING DATE: 1999-04-05  
 ; NUMBER OF SEQ ID NOS: 29  
 ; SOFTWARE: Patent In Ver. 2.0  
 ; SEQ ID NO 12  
 ; LENGTH: 471  
 ; TYPE: PRT  
 ; ORGANISM: Orpinomyces sp. PC-2  
 ; US-09-687-147-12

Query Match 7.9%; Score 175; DB 4; Length 471;  
 Best Local Similarity 24.9%; Pred. No. 2.1e-08;  
 Matches 83; Conservative 45; Mismatches 137; Indels 68; Gaps 16;  
 QY 12 GIDFGCDIDSCPTDTSSVPLLSYKGGDGA-----GOMKHFVE-----DDGLNVFRISA 60  
 Db 35 GMSLGNLTLDASC-----VETLANSKDOTASETCMGVNTTQELYYKLSDGENTFRIPT 88  
 QY 61 TWQFVLNNTYDGLKDELINMGSYNKVYVNAACLETGAYCMIDMHNFRYNGGIIIGGGVSDI 120  
 Db 89 TWSGHFDADPDYKISDVMMKRHVHVVYALNTGAYAILNIHH-ETWVYAFQKNLESARKI 147  
 QY 121 FVDLWVOIAKYVEDNDK-IIFGLNNEPHDL-DIEMWOTQO-----VVTAIRK 167  
 Db 148 LVAIWKQIAAEFGDYDHLIFEGNNEPRKVGDPPEWGTGGDEGNFVNEKMALEFVKTR 207  
 QY 168 AGA--TSOMILLPGTNPASVETVSTGSAEALGKITNPDGSTDLLYEDVHKYLDIN--N 222  
 Db 208 TGGNNARHLMIP-----TYAASVNDGGINNFKYPNNGD-DKVIYSLHSYSPNFALN 258  
 QY 223 SGSHAETTDNVDAFND--FADWLROKROAIISSETGASMEPSCMTAFCAONKAISESD 280  
 Db 259 NGPGA-----ISNFYGENEIDWVMNTINSSFSISKG---IPVILIGEFAVAMNR--DNED 305  
 QY 281 VYIGFVGAGSFPDTSYLLTLPKPKPGNTYDN 313  
 Db 306 -----DREMOEYIKKATALGIPCVIMDN 330

RESULT 12  
 US-09-118-324-2  
 ; Sequence 2, Application US/09118324  
 ; Patent No. 6110720

; GENERAL INFORMATION:  
 ; APPLICANT: Li, Xin-Liang  
 ; APPLICANT: Ljungdahl, Lars G.  
 ; APPLICANT: Chen, Huizhong  
 ; TITLE OF INVENTION: Orpinomyces Cellulase Cele Protein and Coding Sequences  
 ; FILE REFERENCE: 32-98sequence listing  
 ; CURRENT APPLICATION NUMBER: US/09/118,324  
 ; CURRENT FILING DATE: 1998-07-17  
 ; NUMBER OF SEQ ID NOS: 5  
 ; SOFTWARE: Patent In Ver. 2.0  
 ; SEQ ID NO 2  
 ; LENGTH: 477  
 ; TYPE: PRT  
 ; ORGANISM: Orpinomyces sp. PC-2  
 ; US-09-118-324-2

Query Match 7.4%; Score 163.5; DB 3; Length 477;  
 Best Local Similarity 23.8%; Pred. No. 2.8e-07;  
 Matches 71; Conservative 45; Mismatches 107; Indels 75; Gaps 13;  
 QY 49 EDDLNVFRISATWQFVLNNTVDGKDELINMGSYNKVYVNAACLETGAYCMIDMHNFRYNG 108  
 Db 78 KNOGFNFRITTTGHRNGPDKISDVMMKRHVHVVYALNTGAYAILNIHH-EMWNY 136  
 QY 109 GIIGGGVSDI FVDLWVOIAKYVEDNDK-IIFGLNNEPHDL----- 150  
 Db 137 AFSNNLQAKPILAIWKQIAAEFANDVHILFEGNNEPRKVGDPPEWGTGGDEGNFVNEKMALEFVKTR 196  
 QY 151 --IEMWOTQOVTATRKAGATGOM--TLPG-----TNFASVETVSTGSAEALGKITN 202  
 Db 197 EMNAVFLOT-----VNASGNNAIRLMIPTVAAACVNDALDSYR-----KEPT 241  
 QY 203 PDGSTDLLYEDVHKYLDIN--NSGSHAET--TDVND-AFNPFADWLROKROAIIS 255  
 Db 242 NDNK-----VIASVHSYVPNPNLNTGAGAEKTFGSTDIEMANNIKRPLVVRNIPVILIGE 298  
 QY 256 TGASMEPSCMTAFCAONKAISESDVYIGFVGAGSFPDTSYLLTLPKPKPGNTYDN 313  
 Db 299 FGAM-----NRDNESERARMA-----EYIKSATPMGVPCVIMDN 333

RESULT 13  
 US-09-419-459-2  
 ; Sequence 2, Application US/09419459  
 ; Patent No. 6222028  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Liu, Jin-Hao  
 ; APPLICANT: Cheng, Kuo-Juan  
 ; APPLICANT: Tsal, Cheng-Fang  
 ; APPLICANT: Chang, Chia-Chieh  
 ; TITLE OF INVENTION: CELLULOSE ENZYMES  
 ; FILE REFERENCE: 08919/037001  
 ; CURRENT APPLICATION NUMBER: US/09/419,459  
 ; CURRENT FILING DATE: 1999-10-15  
 ; NUMBER OF SEQ ID NOS: 10  
 ; SOFTWARE: FastSeq for Windows Version 4.0  
 ; SEQ ID NO 2  
 ; LENGTH: 582  
 ; TYPE: PRT  
 ; ORGANISM: Piromyces rhizinflata  
 ; US-09-419-459-2

Query Match 7.3%; Score 162.5; DB 4; Length 582;  
 Best Local Similarity 19.9%; Pred. No. 4.7e-07;  
 Matches 81; Conservative 73; Mismatches 174; Indels 79; Gaps 15;  
 QY 10 IPGIDFGCDIDSCPTDTSSVPLLSYKGGDAGQ-----MKHFAEDGLNVFRI 58  
 Db 147 IKEMNFGNLCNT--MDAQCIEYLNTERKQDPASETCWCNPKTDEMDKVLIDNQFNVERI 204  
 QY 59 SATWQFVLNNTVDGKDELINMGSYNKVYVNAACLETGAYCMIDMHNFRYNGGIIIGGGVSD 118  
 Db 205 PTTWSGHFGEAPDYKIDKMLKRVHEVVDYPRKNGAFVILNLH-ETWVNAFSETLDTAK 263

```
US-09-687-147-15
; Sequence 15, Application US/09687147
; Patent No. 6268198
; GENERAL INFORMATION:
; APPLICANT: Li, Xin-Liang
; APPLICANT: Ljungdahl, Lars G.
; APPLICANT: Chen, Huizhong
; TITLE OF INVENTION: Cellulases and Coding Sequences
; FILE REFERENCE: 42-96a
; CURRENT APPLICATION NUMBER: US/09/687,147
; CURRENT FILING DATE: 2000-10-12
; PRIOR APPLICATION NUMBER: US 60/027,883
; PRIOR FILING DATE: 1996-10-04
; PRIOR APPLICATION NUMBER: PCT US97/18008
; PRIOR FILING DATE: 1997-10-03
; PRIOR APPLICATION NUMBER: 09/286,691
; PRIOR FILING DATE: 1999-04-05
; NUMBER OF SEQ ID NOS: 29
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 15
; TYPE: PRF
; LENGTH: 473
; ORGANISM: Neocallimastix patriciarum
US-09-687-147-15

Query Match      7.9%; Score 175.5; DB 4; Length 473;
Best Local Similarity 22.2%; Pred. No. 1.9e-08;
Matches 88; Conservative 55; Mismatches 172; Indels 81; Gaps 17;

QY 12 GIDFGCDIDGSCPTDTSVPLSYKGGDGA-----GOMKHFAE-----DDGLNVRISA 60
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 35 GWSLGNLTLDASC-----FETLDYKRNQIASSETCNGNVTTOLELYKLSLGLGNTFRIPY 88
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 61 TMOFVLNNTVDGKLDELINMGSYKYNACLETGAYCMIDHNFARYNGGIIGGGVSDDI 120
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 89 TWSGHNADPKYKIDVMKMRVHEVDYALNTGVALINLHH-ETWNHAFOKULESAKKI 147
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 121 FVDLWQIAKYEDNDK-IIFGLMNEPHDL-----DIEIWT-----AOTQKRVYTAIRK 167
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 148 LVAIMQIAAEFGDYDEHLIFEGMNEPRKVGDAEMNGGYEGMNVNEMNDLFVYTIIRA 207
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 168 AGATSGM--ILPGTNPASVETVSTGSAALCKITNPDGSDTLDFYVHKYLDIN---N 222
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 208 TGGNNALRHLMP-----TYAACINDGAINNFKFPGSD--KRYVYLSHSYSPYFALN 258
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 223 SGSHACCTDNDVAFND--FADWLRONKROAIISETGASMEPCMTAFCAONKAISENSD 280
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 259 NGAGA-----ISNFDGSEIDMAANTINSKFIISRG-----IPYIGFEGAMNNEDDRE 308
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 281 VYIGFVGWAGSFDTSYITLTPPLKRP-----GNYTDNKLMLNCEILDQFTLDEKY----- 330
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 309 RMAEY-----YIKKATSIGVPCVIMWNGYFEGEGEGRGLNRSRLOVYVYPRLVN 357
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 331 ---RPTPTISITAEETATATATSDGDAPSTYPIF 363
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 358 GLIKGLGNSIKRTYIRRTYTTTSSQSOPITNDSCF 399
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |

RESULT 9
US-09-118-324-5
; Sequence 5, Application US/09118324
; Patent No. 6110720
; GENERAL INFORMATION:
; APPLICANT: Li, Xin-Liang
; APPLICANT: Ljungdahl, Lars G.
; APPLICANT: Chen, Huizhong
; TITLE OF INVENTION: Orpinomyces Cellulase CelE Protein and Coding Sequences
; FILE REFERENCE: 32-98sequence listing
; CURRENT APPLICATION NUMBER: US/09/118,324
; CURRENT FILING DATE: 1998-07-17
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 5
```

```
; LENGTH: 471
; TYPE: PRF
; ORGANISM: Orpinomyces sp. PC-2
US-09-118-324-5

Query Match      7.9%; Score 175; DB 3; Length 471;
Best Local Similarity 24.9%; Pred. No. 2.1e-08;
Matches 83; Conservative 45; Mismatches 137; Indels 68; Gaps 16;

QY 12 GIDFGCDIDGSCPTDTSVPLSYKGGDGA-----GOMKHFAE-----DDGLNVRISA 60
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 35 GWSLGNLTLDASC-----VETLWYSKDQTASETCNGNVTTOLELYKLSLGLGNTFRIPY 88
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 61 TMOFVLNNTVDGKLDELINMGSYKYNACLETGAYCMIDHNFARYNGGIIGGGVSDDI 120
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 89 TWSGHNADPKYKIDVMKMRVHEVDYALNTGVALINLHH-ETWNHAFOKULESAKKI 147
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 121 FVDLWQIAKYEDNDK-IIFGLMNEPHDL-DIEIWAQTCQK-----VYTAIRK 167
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 148 LVAIMQIAAEFGDYDEHLIFEGMNEPRKVGDAEWYTGDDQEGMNVNEMNDLFVYTIIRA 207
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 168 AGA--TSQMLLPGTNPASVETVSTGSAALCKITNPDGSDTLDFYVHKYLDIN---N 222
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 208 TGGNNALRHLMP-----TYAASVNDGSIINFKFPGSD--KRYVYLSHSYSPYFALN 258
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 223 SGSHACCTDNDVAFND--FADWLRONKROAIISETGASMEPCMTAFCAONKAISENSD 280
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 259 NGAGA-----ISNFDGSEIDMAANTINSKFIISRG-----IPYIGFEGAMNNEDDRE 305
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 281 VYIGFVGWAGSFDTSYITLTPPLKRPNGYTDN 313
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 306 -----DREMQEYIYKKAATAGIIPCVIWDN 330
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |

RESULT 10
US-09-286-691-12
; Sequence 12, Application US/09286691
; Patent No. 6190189
; GENERAL INFORMATION:
; APPLICANT: Li, Xin-Liang
; APPLICANT: Ljungdahl, Lars G.
; APPLICANT: Chen, Huizhong
; TITLE OF INVENTION: Cellulases and Coding Sequences
; FILE REFERENCE: 42-96
; CURRENT APPLICATION NUMBER: US/09/286,691
; CURRENT FILING DATE: 1999-04-05
; EARLIER APPLICATION NUMBER: US 60/027,883
; EARLIER FILING DATE: 1996-10-04
; EARLIER APPLICATION NUMBER: PCT US97/18008
; EARLIER FILING DATE: 1997-10-03
; NUMBER OF SEQ ID NOS: 29
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 12
; LENGTH: 471
; TYPE: PRF
; ORGANISM: Orpinomyces sp. PC-2
US-09-286-691-12

Query Match      7.9%; Score 175; DB 4; Length 471;
Best Local Similarity 24.9%; Pred. No. 2.1e-08;
Matches 83; Conservative 45; Mismatches 137; Indels 68; Gaps 16;

QY 12 GIDFGCDIDGSCPTDTSVPLSYKGGDGA-----GOMKHFAE-----DDGLNVRISA 60
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 35 GWSLGNLTLDASC-----VETLWYSKDQTASETCNGNVTTOLELYKLSLGLGNTFRIPY 88
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 61 TMOFVLNNTVDGKLDELINMGSYKYNACLETGAYCMIDHNFARYNGGIIGGGVSDDI 120
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 89 TWSGHNADPKYKIDVMKMRVHEVDYALNTGVALINLHH-ETWNHAFOKULESAKKI 147
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 121 FVDLWQIAKYEDNDK-IIFGLMNEPHDL-DIEIWAQTCQK-----VYTAIRK 167
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 148 LVAIMQIAAEFGDYDEHLIFEGMNEPRKVGDAEWYTGDDQEGMNVNEMNDLFVYTIIRA 207
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
```

OY 237 FNDPDLRQKROAIISGTGASMPSCMTAFCAQNAKAISENSDYITGVGAGSPTDS 296  
 DB 248 ITDQWMLKDKKKGFISETYAGSNDVCRSAVSGMLEYMANNTDVKGASWMAAGPMWGD 307  
 OY 297 YILTLTP 303  
 DB 308 YIFSLP 314

# RESULT 6 US-09-478-816-4

Sequence 4, Application US/09478816  
 Patent No. 6306635  
 GENERAL INFORMATION:  
 APPLICANT: VAN DEN BROECK, HENRIETTE C.  
 APPLICANT: DE GRAAF, LEENDERT H.  
 APPLICANT: VISSER, JACOB  
 APPLICANT: VAN OYEN, ALBERT J. J.  
 TITLE OF INVENTION: FUNGAL CELLULASES  
 NUMBER OF SEQUENCES: 4  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: MORRISON & FOERSTER  
 STREET: 2000 Pennsylvania Avenue, NW, suite 5500  
 CITY: Washington  
 STATE: DC  
 COUNTRY: USA  
 ZIP: 20006-1888  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Diskette  
 OPERATING SYSTEM: DOS  
 SOFTWARE: FastSeq for Windows Version 2.0  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/09/478,816  
 FILING DATE:  
 CLASSIFICATION:  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US/08/849,751  
 FILING DATE: 12-JUN-1997  
 APPLICATION NUMBER: PCT/EP96/04511  
 FILING DATE: 14-OCT-1996  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Murashige, Kate H  
 REGISTRATION NUMBER: 29,959  
 REFERENCE/DOCKET NUMBER: 24615-20094.00  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 202-887-1500  
 TELEFAX: 202-822-0168  
 TELEX:  
 INFORMATION FOR SEQ ID NO: 4:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 331 amino acids  
 TYPE: amino acid  
 STRANDEDNESS: single  
 TOPOLOGY: linear  
 MOLECULE TYPE: protein  
 FRAGMENT TYPE: Internal  
 US-09-478-816-4

Query Match 16.3%; Score 361; DB 4; Length 331;  
 Best Local Similarity 29.0%; Pred. No. 1.3e-26;  
 Matches 89; Conservative 56; Mismatches 130; Indels 32; Gaps 7;

OY 4 KYLGVALPGIDFGCDIDGSCPT-----DTSSVPLLSYKGDGAGOMKHPAEDGILANFRI 58  
 DB 33 EMGGSNSGAEFGTINIGVGTDIYFPDPSTISTLLK-----GMNFRV 77  
 OY 59 SATWQFLVNTYDGLDELWMSYKRVNACLETGAYCMIDMNFARYNGIIGGGSVD 118  
 DB 78 QPMERLLPSSMTGSIDEEXLANLTVVKAVTGGAHALLDPNHYGNEIIS---ST 133  
 OY 119 DIFVDLWQIAKYEDNDKIIIFGLMNEPHDLIEIWAQTCQKVVTIIRKAGATSOMLLP 178

DB 134 SDFQTFQNLGQYKNDLVNEDTNNREYDMQDVLNLQNAINGIRAGA-SQYIFE 192  
 OY 179 GTFASVEYVYSTGSAEALGKTNPDGSTDLIEPVHYLIDNNSGIAECTDNY--DA 236  
 DB 193 GNSWTGANTWVDN--DNKKILTDE---DKIVEMHOLYSDSGTSETCVSGTIGER 247  
 OY 237 FNDPDLRQKROAIISGTGASMPSCMTAFCAQNAKAISENSDYITGVGAGSPTDS 296  
 DB 248 ITDQWMLKDKKKGFISETYAGSNDVCRSAVSGMLEYMANNTDVKGASWMAAGPMWGD 307  
 OY 297 YILTLTP 303  
 DB 308 YIFSLP 314

## RESULT 7

US-09-286-691-15  
 Sequence 15, Application US/09286691  
 Patent No. 6190189  
 GENERAL INFORMATION:  
 APPLICANT: Li, Xin-Liang  
 APPLICANT: Ljungdahl, Lars G.  
 APPLICANT: Chen, Huizhong  
 TITLE OF INVENTION: Cellulases and Coding Sequences  
 FILE REFERENCE: 42-96  
 CURRENT APPLICATION NUMBER: US/09/286,691  
 CURRENT FILING DATE: 1999-04-05  
 EARLIER APPLICATION NUMBER: US 60/027,883  
 EARLIER FILING DATE: 1996-10-04  
 EARLIER APPLICATION NUMBER: PCT US97/18008  
 EARLIER FILING DATE: 1997-10-03  
 NUMBER OF SEQ ID NOS: 29  
 SOFTWARE: PatentIn Ver. 2.0  
 SEQ ID NO 15  
 LENGTH: 473  
 TYPE: PRT  
 ORGANISM: Neocallimastix patriciarum  
 US-09-286-691-15

Query Match 7.9%; Score 175.5; DB 4; Length 473;  
 Best Local Similarity 22.2%; Pred. No. 1.9e-08;  
 Matches 88; Conservative 55; Mismatches 172; Indels 81; Gaps 17;

OY 12 GIDFGCDIDGSCPTDTSSVPLLSYKGDG-----GOMKHPAE-----DDGLNVFRTSA 60  
 DB 35 GWSLGNLTLDATC-----FETLDYNNKQIASETCMGNVKTQELXYKLSDLGFTFRPT 88  
 OY 61 TWQFVLNNTYDGLDELWMSYKRVNACLETGAYCMIDMNFARYNGIIGGGSVDI 120  
 DB 89 TWSGHFGNAPDYKINDOMKRVHIEIVYAIINTGYAILNIHH-ETWNHAFQKNLESARKI 147  
 OY 121 FVDLWQIAKYEDNDKIIIFGLMNEPHDL-----DLEIWAQTCQKVVTIIRK 167  
 DB 148 LVAIWKQIAAEFADYDELIFEGNNEPRKAGDPAENMGDYEENVEYNNMDLFVKITRA 207  
 OY 168 AGATSOV--ILLPCTNFASVETVSTGSAEALGKITPNPDGSTDLIEPVHYLIDNN 222  
 DB 208 TGGNNALRHLMP-----TYAACINDGAINNFKFPSCD-DKVIYLSLHISYSPYNALN 258  
 OY 223 SGSAEETTDNDVAFNDPA--DWLRQKROAIISGTGASMPSCMTAFCAQNAKAISENSD 280  
 DB 259 NGAGA-----ISNFYDGESEIDWAMNTINSKFISRG---IPVILGFGAMNRRNEDRE 308  
 OY 281 VYIGFVWGAGSPTDSYILTLPLGKP-----GNYTDNKLAMNECIIIDQFTLDEKY----- 330  
 DB 309 RMAEY-----YIKATSIIGVPCVIWMDGYEGEGERGLINRSTLQVYVPLVN 357  
 OY 331 ---RPPTSIATAEETATATATSDGAPSTYPIF 363  
 DB 358 GLIKGLGNSIKTFTTIRRTTTTTTSSQSQPINNDSCF 393

RESULT 8



Db 393 ILTETPTSGNSMTDTSILVSSCL 415

RESULT 2

US-08-169-948B-16

Sequence 16, Application US/08169948B

Patent No. 5861271

GENERAL INFORMATION:

APPLICANT: Fowler, Timothy

APPLICANT: Ward, Michael

APPLICANT: Clarkson, Kathleen

APPLICANT: Collier, Katherine

APPLICANT: Latenas, Edmund

TITLE OF INVENTION: No. 5861271el Cellulase, Enzymes and Systems

TITLE OF INVENTION: For Their Expression

NUMBER OF SEQUENCES: 48

CORRESPONDENCE ADDRESSES:

ADDRESSEE: Genencor International

STREET: 180 Kimball Way

CITY: South San Francisco

STATE: CA

COUNTRY: USA

ZIP: 94080

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/169,948B

FILING DATE: DEC 17 1993

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Horn, Margaret A.

REGISTRATION NUMBER: 33,401

REFERENCE/DOCKET NUMBER: GC226

TELECOMMUNICATION INFORMATION:

TELEPHONE: (415) 742-7536

TELEFAX: (415) 742-7217

INFORMATION FOR SEQ ID NO: 16:

SEQUENCE CHARACTERISTICS:

LENGTH: 327 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-169-948B-16

Query Match 42.6%; Score 942.5; DB 2; Length 327;  
Best Local Similarity 51.7%; Pred. No. 9.8e-83;  
Matches 167; Conservative 61; Mismatches 90; Indels 5; Gaps 2;

3 IKYLGVAIPGIDRCDDIDGSCPTDTSVPLSTYGG----DGACOMKHFADDDLNVFRI 58  
2 VRFAGVNIAGFDECTDGTCTVTSKYVPLKFNFGSNYPDGIOMOFVNEDEMTIFRL 61  
59 SATQFVNLNTVDKLDLNLNMGSYNKVYNACTLETGAYCMIDMNFARYNGIIOGGVSD 118  
62 PVGQIYLVNNGNLNLDSTISKIDLVQGLSGAYCIVDINHVARNGIIOGGGPTN 121  
119 DIPVDLWVQIAKYEDNDKIIFGLMNEPHDIDIEIWAOTCOKVYTAIRKAGATSOMILLP 178  
122 AOFSLMSQLASKYASOSRWVFGIMNEPHDVNIMTMAATVOEVYTAIRNAGATSOFLSLP 181  
179 GTNFASVETVYSGAELIGKITNPDGSTDLYFDVHKYLDINNNGSHAECTTNVD-AF 237  
182 GNDQSGAFLSDGSAALSOVTNPDSITNLFVDVHKYLDSDNSGTHAECTTNNDGAF 241  
238 NDFADWLNRONKROAIISSETGASMEPSCMTAFCAONKAISENSDVIYIGFVGAGASFPDTSY 297  
242 SPLATWLRONNRKRAILLETGGGNVOSCIODMCOQIOYLONNSDVIYIGFVGAGASFPDTSY 301  
298 ILTLPGLKPGNYTDNKLINNECI 320

Db 302 VLTEPTSSGNSMTDTSILVSSCL 324

RESULT 3

US-08-448-873-16

Sequence 16, Application US/08448873

Patent No. 5874276

GENERAL INFORMATION:

APPLICANT: Fowler, Timothy

APPLICANT: Ward, Michael

APPLICANT: Clarkson, Kathleen

APPLICANT: Collier, Katherine A.

TITLE OF INVENTION: No. 5874276el Cellulase Enzymes and Systems

TITLE OF INVENTION: For Their Expressions

NUMBER OF SEQUENCES: 48

CORRESPONDENCE ADDRESSES:

ADDRESSEE: Genencor International

STREET: 180 Kimball Way

CITY: South San Francisco

STATE: CA

COUNTRY: USA

ZIP: 94080

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/448,873

FILING DATE:

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/169,948

FILING DATE: 17-DEC-1993

ATTORNEY/AGENT INFORMATION:

NAME: Stone, Christopher L.

REGISTRATION NUMBER: 35,686

REFERENCE/DOCKET NUMBER: GC226D14

TELECOMMUNICATION INFORMATION:

TELEPHONE: (415) 742-7555

TELEFAX: (415) 742-7217

INFORMATION FOR SEQ ID NO: 16:

SEQUENCE CHARACTERISTICS:

LENGTH: 327 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-448-873-16

Query Match 42.6%; Score 942.5; DB 2; Length 327;  
Best Local Similarity 51.7%; Pred. No. 9.8e-83;  
Matches 167; Conservative 61; Mismatches 90; Indels 5; Gaps 2;

3 IKYLGVAIPGIDRCDDIDGSCPTDTSVPLSTYGG----DGACOMKHFADDDLNVFRI 58  
2 VRFAGVNIAGFDECTDGTCTVTSKYVPLKFNFGSNYPDGIOMOFVNEDEMTIFRL 61  
59 SATQFVNLNTVDKLDLNLNMGSYNKVYNACTLETGAYCMIDMNFARYNGIIOGGVSD 118  
62 PVGQIYLVNNGNLNLDSTISKIDLVQGLSGAYCIVDINHVARNGIIOGGGPTN 121  
119 DIPVDLWVQIAKYEDNDKIIFGLMNEPHDIDIEIWAOTCOKVYTAIRKAGATSOMILLP 178  
122 AOFSLMSQLASKYASOSRWVFGIMNEPHDVNIMTMAATVOEVYTAIRNAGATSOFLSLP 181  
179 GTNFASVETVYSGAELIGKITNPDGSTDLYFDVHKYLDINNNGSHAECTTNVD-AF 237  
182 GNDQSGAFLSDGSAALSOVTNPDSITNLFVDVHKYLDSDNSGTHAECTTNNDGAF 241  
238 NDFADWLNRONKROAIISSETGASMEPSCMTAFCAONKAISENSDVIYIGFVGAGASFPDTSY 297  
242 SPLATWLRONNRKRAILLETGGGNVOSCIODMCOQIOYLONNSDVIYIGFVGAGASFPDTSY 301





CC and a method for producing recombinant cellulase by culturing these  
CC host cells.

XX Sequence 471 AA:

Query Match 7.9%; Score 175; DB 19; Length 471;  
Best Local Similarity 24.9%; Pred. No. 4.1e-07;  
Matches 83; Conservative 45; Mismatches 137; Indels 68; Gaps 16;

QY 12 GIDFGCDINGSCPTDSSVPLSLYKGGDGA-----GQKHFAE-----DDGLNVRISA 60  
DB 35 GMSLGTTLDASC-----VETLNYSKDQFASETCWNVKTQDELKYKLSDLGNTFRIPR 88  
QY 61 TQGFVNTVNTVDGKLDLNMGSYKNVYNACTETGAYCMIDMHNFRYNGIIGGGVSDDI 120  
DB 89 TMSGHGCDAPDKISDVMMKRVHEVVDYALNTGVAIILNIH-ETMNYAFQKLESAAKI 147  
QY 121 FVDLWQIAKYYEDNDK-IIFGLMNEPHDL-DIEIWAQTCQK-----VVTAIRK 167  
DB 148 LVAIMKQIAAEFGDYDEHLIFEGMNEPRKYGDAEWGTGQEGMNFVNEKNALFVKIRA 207  
QY 168 AGA--TSCMILLPGTFASVETVSTGSAEALGKITNPGOSTDLKFDVHKYLDIN---N 222  
DB 208 TCGNNANRHLMP-----TYAASVNDGSIINFRYPNGD-DKVIYSLHSYSPYFALN 258  
QY 223 SGSHAECTDNDVAFND--FADWLRONKROAISETGASMEPSCMTAFCAQNKAISENSD 280  
DB 259 NPGCA-----ISNFYDGNIEDVMTNTINSFISKG-----IPYIIEFVAMNR---DNE 305  
QY 281 VIIEGVGAGSFDTSYILTLPLGKPGNYTDN 313  
DB 306 -----DREMOEYIKKATFALGIPCVIMDN 330

RESULT 14

AAB08774 AAB08774 standard; Protein; 477 AA.

XX AAB08774;

DT 02-JAN-2001 (first entry)

XX Amino acid sequence of a CelE cellulase polypeptide.  
DE CelE cellulase; endoglucanase; carboxymethylcellulose; beta-glucan;  
KW lichenin; paranitrophenyl-beta-cellulobioside; cellulobiose.

XX Orpinomyces sp.

OS US6110720-A.

PN 29-AUG-2000.

PD 17-JUL-1998; 98US-0118324.

PF 17-JUL-1998; 98US-0118324.

PR 17-JUL-1998; 98US-0118324.

PA (UYGE-) UNIV GEORGIA RES FOUND INC.

PI Li X, Chen H, Ljungdahl LG;

DR WPI; 2000-586355/55.

N-PSDB; AAA64503.

PT Novel nucleic acids encoding cellulase from Orpinomyces species PC-2  
XX which produces cellulobiose from carboxymethylcellulose, cellotetraose,  
XX beta-glucan and lichenin

PS Claim 2; Column 9-12; 14pp; English.

CC The present sequence represents a CelE cellulase. The polypeptide has  
CC endoglucanase activity, and hydrolyzes carboxymethylcellulose, to  
CC beta-glucan, lichenin and paranitrophenyl-beta-cellulobioside, to

CC produce cellulobiose. The polypeptide is derived from Orpinomyces sp.  
CC PC-2, and shows homology to several other endoglucanases. The CelE  
CC polypeptide has a catalytic domain to glycosyl hydrolases of Family 5  
CC anaerobic bacteria. The CelE gene is devoid of introns.

XX Sequence 477 AA:

Query Match 7.4%; Score 163.5; DB 21; Length 477;  
Best Local Similarity 23.8%; Pred. No. 4.8e-06;  
Matches 71; Conservative 45; Mismatches 107; Indels 75; Gaps 13;

QY 49 EDDGLNVRISATWQFVLTNTVDGKLDLNMGSYKNVYNACTETGAYCMIDMHNFRYNG 108  
DB 78 KNGFNVFRIPTWTGHEFGNGPDYKISDVMMRVHEVVDYALNTGVAIILNIH-EMWNY 136  
QY 109 GIIGGGVSDIFVDLWQIAKYYEDNDK-IIFGLMNEPHDL----- 150  
DB 137 AFSNNLQKAKPILAAIMKQIAAEFANYDEHLIFEGMNEPRKYGDAEWGTGQEGMNFV 196  
QY 151 --TEIWAQTCQKVTWTAIRKAGATSON--ILLPG---TTFASVETVSTGSAEALGKITN 202  
DB 197 ENNAVFLQT-----VRASGNNAIRHLMIPTFYAACVNDGALESYR-----KFP 241  
QY 203 PDGSTDLKFDVHKYLDIN---NSGSHAECT--TDND-APNDPADWLRONKROAIISE 255  
DB 242 NDKK---VIASVHSYVPYNFALNTGAGAEKTFGSTSDIEWAMNNIKRFLVDNRNIPYI 298  
QY 256 TGSMEPSCMTAFCAQNKAISENSDVIIGFVGAGSFDTSYILTLPLGKPGNYTDN 313  
DB 299 FGAM-----NRDNESERARMA-----EYIKSATAMGVPCVIMDN 333

RESULT 15

AAB62467 AAB62467 standard; Protein; 582 AA.

XX AAB62467;

DT 09-JUL-2001 (first entry)

DE P. rhizinflata egIIA protein partial sequence.

XX Cellulase; egIIA; fungus; polysaccharide; glycosidic linkage; enzyme;  
KW animal feed; alcohol; fermentation.

XX Pithomyces rhizinflata.

OS Key Location/Qualifiers

FT 1..109 Region /note="repeat region (AAB62470)"

FT 110..499 Domain /note="catalytic domain (AAB62468)"

FT Region /note="repeat region (AAB62471)"

PN US6222028-B1.

PD 24-APR-2001.

PE 15-OCT-1999; 99US-0419459.

PR 15-OCT-1999; 99US-0419459.

PA (SINT-) ACAD SINICA.

PI Liu J, Cheng K, Tsai C, Chang C;

DR WPI; 2001-307598/32.

N-PSDB; AAF83292.

CC New isolated nucleic acid encoding a cellulase enzyme of fungus  
PT Pithomyces rhizinflata, for producing cellulase that hydrolyzes  
PT polysaccharides containing beta-1,3 or beta-1,4 glycosidic linkages

Db 193 GNSWTGAMTWVDV--DNMKNLTPDPE---DKIYVEMHOYLDSGSGTSEFCVSFTIGKER 247  
 QY 237 FNDPADLRLONKROAIISFTGASMEPCMTAFCAONKAISENSDVIYIGFVGWAGSFDTS 296  
 Db 248 ITDPTQWLKKNKVKVGFGEYAGSGNDVCRSAVSGMLEYMANNTDVMKASWMAAGPRMWD 307  
 QY 297 YIILTP 303  
 Db 308 YIFSLP 314

## RESULT 12

AA047496  
 ID AA047496 standard; Protein: 800 AA.

AA047496:

21-JUL-1994 (first entry)

Translated sequence of domains I and II of celd cDNA in clone PCNP4.

Cellulase; celd; PCNP4; anaerobic rumen.

Neocallimastix patriciarum.

Key Location/Qualifiers

Region 1..37 /label= N-terminus of beta-galactosidase alpha-peptide

Misc-difference 38..42 /note= "derived from 5' oligo linker"

WO9400578-A.

06-JAN-1994.

24-JUN-1993: 93MO-AU00307.

24-JUN-1992: 92AU-0003096.

(CSIR) COMMONWEALTH SCI & IND RES ORG.

Aylward JH, Gobius KS, Orpin CG, Xue GP;

WPI: 1994-026214/03.

N-PSDB; AA055036.

Cloning of cellulase clones from anaerobic rumen - by isolating

mRNA from culture, converting to cDNA etc. fungi, producing

enzymes useful in food processing etc., and DNA for modifying

rumen or silage bacteria

Claim 10: Page 41-45; 71pp; English.

Clone PCNP4.1 encodes celd. It was derived as follows. N.

patriciarum was used to construct a cDNA library in ZAPIT (in E.

coli). Transformants were selected for enzyme activity using

selective media. 11 colonies were positive, and of these 10 had the

same restriction pattern, and the longest of these was designated

celd (PCNP4.1)(AA055036)(AA047496.R49102). A similar process was used

to isolate the xylanase clone PXN-rac (AA055037.R47497). An enzyme

composition contg. celd and xylanase is claimed.

Sequence 800 AA;

Query Match 8.3%; Score 184; DB 15; Length 800;

Best Local Similarity 20.1%; Pred. No. 1.4e-07;

Matches 92; Conservative 70; Mismatches 193; Indels 102; Gaps 17;

QY 12 GIDFGCDIDSCPTDSSVPLSYKGGDAGQ-----MKHAEEDGLNWFRIISA 60

Db 59 GNMIGNTLDAQ-----IEVLNVDKQDTASERCWGNPKTTEDMFXYLMNQFNVRPIPT 112

QY 61 TWQFVLANTVDGKIDELNMGSYKNVNAACLETGAYCMIDMNFARYNGIIGGGSDDI 120  
 Db 113 TWSGHFEADPDYKINEMKLRVHEIVDYPKNGAFVILNLHH-ETWNHASETLDIAKEI 171  
 QY 121 FVDLWVOIAKYVEDNDK-ITFGLINEPHDIDIE-WAQTQCK-----VTAIRK 167  
 Db 172 LEKISQIAKEFKDYDEHLIFEGLINEPRKNDTPVEMTGCDGDAVANANAVFKITRS 231  
 QY 168 AGATS--QMILLPGTNPASVETVYSTGSAEALGKITPDGSTDLLYDVKRYLIDNNGS 225  
 Db 232 SGNNPKRHLIMP-----PYAACNENSEFKNFIPE-DDDKVIASVHAAYAPYALN 282  
 QY 226 HAECTTDNDAF-NDFADLRLONKROAIISFTGASMEPCMTAFCAONKAISENSDVIYIG 284  
 Db 283 NGEAGAVKFPDAAKGRKDEEMNINLMKKRFRVQG-----IPMILGEYGANRDEEDRAAMAE 338  
 QY 285 FVGWAGSFPDSYILTLPLGKPGNYTDNKLMBNC-----ILDOFTLDEKRPPTISIT 339  
 Db 339 F-----YMEKTAAGVPOVWMDNGIFECTGERFGLDKKLNKIY---PTIYVA 384  
 QY 340 AAE-----ETATATATSDGAPSTTKPIFEETASP-TPNAVTKPSPDSDSD----- 387  
 Db 385 LQKRGLEVNVAHIEKKPEPTKTTEPEPTETTSPEEPAETTNPEPTGNIRDISKE 444  
 QY 388 -----DCKDAASMSAOG 400  
 Db 445 LIKEMNFGMNLGNTLDAQIEYLVNDRDQTASEFCWG 481

## RESULT 13

AAW56742  
 ID AAW56742 standard; Protein: 471 AA.

AAW56742:

14-SEP-1998 (first entry)

Orpinomyces cellulase celdB.

Cellulase; endoglucanase; celllobiondrolase; celdB.

Orpinomyces sp. strain PC-2.

WO9814597-A1.

09-APR-1998.

03-OCT-1997: 97WO-US18008.

04-OCT-1996: 96US-0027883.

(UYGE-) UNIV GEORGIA RES FOUND INC.

Chen H, Li X, Ljungdahl LG;

WPI: 1998-240096/21.

N-PSDB; AAW29477.

New recombinant DNA encoding Orpinomyces cellulase protein - useful

for, e.g. producing recombinant Orpinomyces cellulase in host cell

Claim 1; Page 38-40; 69pp; English.

This polypeptide comprises cellulase celdB of Orpinomyces sp. strain

OC-2, an anaerobic bovine rumen fungus Orpinomyces sp. PC-2. Its

amino acid sequence was deduced from an isolated cDNA clone (see

AAV29477). celdB has cellobiohydrolase activity, with highest

activity at pH 5.2-6.2 and 50 degC. celdA (see AAW56738) and celdC

(see AAW56739) cellulolytic enzymes of Orpinomyces sp. PC-2 are also

provided. Recombinant DNA molecules encoding Orpinomyces cellulase

proteins are claimed, as well as recombinant cells selected from

Saccharomyces cerevisiae, Escherichia coli, Aspergillus,

Trichoderma reesei, Pichia, Penicillium, Streptomyces or Bacillus,

```

RESULT 10
AAW14596
ID AAW14596 standard; Protein: 331 AA.
XX
AC AAW14596;
XX
DT 22-JUL-1997 (first entry)
XX
DE Aspergillus niger cellulase.
XX
KW Cellulase; protein detection; enzyme detection; library screening.
XX
OS Aspergillus niger strain N400 (CBS120.49).
XX
PN WO9713853-A2.
XX
PD 17-APR-1997.
XX
PF 14-OCT-1996; 96WO-EP04510.
XX
PR 13-OCT-1995; 95EP-0202777.
XX
PA (KONN ) GIST-BROCADES BV.
XX
PI De Graaff LH, Van Den Broeck HC, Van Ooyen AJJ, Vlaser J;
XX
DR MPI; 1997-235889/21.
XX
DR N-PSDB: AAT63043.
XX
PS Identifying DNA fragments encoding proteins, e.g. for new enzyme
PT discovery - by direct screening of a cDNA library in bacteria
PT transformed with DNA from eukaryotic organism producing the protein
XX
XX
XX Claim 12; Page 16-17; 30pp; English.
XX
A cellulase (AAW14596) is encoded by a cDNA clone (AAT63043) derived
CC from Aspergillus niger N400 (CBS120.49). The cDNA clone was
CC isolated using a method for identifying a DNA fragment encoding a
CC protein of interest. The method involves the direct screening of a
CC cDNA library prep. in bacteria (e.g. E. coli) transformed with DNA
CC from a eukaryotic organism (e.g. A. niger) that produces the
CC protein. Screening for cellulase-producing clones can be performed
CC on plates containing CM-cellulase and Congo Red. The method was
CC utilised in the identification of cDNA clones (AAT63042-46) coding
CC for A. niger cellulase, xylanase and arabinoxylan degrading enzymes
CC (AAW14595-99).
XX
SQ Sequence 331 AA:
Query Match 16.3%; Score 361; DB 18; Length 331;
Best Local Similarity 29.0%; Pred. No. 1.9e-24;
Matches 89; Conservative 56; Mismatches 130; Indels 32; Gaps 7;

```

```

OY 297 YITLTP 303
DB 308 YITSLTP 314
RESULT 11
AAW15563
ID AAW15563 standard; Protein: 331 AA.
XX
AC AAW15563;
XX
DT 08-JUL-1997 (first entry)
XX
DE Cellulase polypeptide (cDNA clone 64 product).
XX
KW Cellulase; carboxymethylcellulase; endoglucanase; beta-glucanase;
KW cellulolytic enzyme; wine; beer; juice clarification; paper; pulp;
KW textile; feedstuff.
XX
OS Aspergillus niger strain N400 (CBS120.49).
XX
PN WO9713862-A1.
XX
PD 17-APR-1997.
XX
PF 14-OCT-1996; 96WO-EP04511.
XX
PR 13-OCT-1995; 95EP-0202776.
XX
PA (KONN ) GIST-BROCADES BV.
XX
PI De Graaff LH, Van Den Broeck HC, Van Ooyen AJJ, Vlaser J;
XX
DR MPI; 1997-235898/21.
XX
DR N-PSDB: AAT60427.
XX
PS Aspergillus cellulolytic enzyme - degrades carboxymethyl cellulose,
PT endoglucan and beta-glucan, used e.g. in production of wine, animal
PT feeds etc.
XX
XX
XX Example 2; Page 18-19; 28pp; English.
XX
Novel enzymes (AAW15562 and AAW15563) that have carboxymethylcellulase,
CC endoglucanase and beta-glucanase activities are encoded by cDNA
CC clones 12 and 64 (AAT60426-27) identified by expression cloning in
CC E. coli using a plasmid vector. They can be expressed in yeast,
CC bacteria or fungi hosts and used to degrade cellulose and glycans,
CC esp. in the manufacture of wine, beer or fruit/vegetable juices
CC e.g. to liquefy plant walls and reduce membrane fouling, and in
CC the textile, detergent and paper/pulp industries e.g. to impart a
CC worn appearance to clothing. The cellulases can also be used to
CC improve animal feed utilisation.
XX
SQ Sequence 331 AA:
Query Match 16.3%; Score 361; DB 18; Length 331;
Best Local Similarity 29.0%; Pred. No. 1.9e-24;
Matches 89; Conservative 56; Mismatches 130; Indels 32; Gaps 7;

```

```

OS Aspergillus niger.
XX Key Location/Qualifiers
XX Key misc_difference 6
FT misc_difference 97 /note- "encoded by GCG"
FT misc_difference 97 /note- "encoded by TAG"
FT misc_difference 130 /note- "encoded by GAC"
XX
XX MO9806858-A1.
XX
XX 19-FEB-1998.
XX
XX 11-AUG-1997; 97WO-EP04415.
XX
XX 15-AUG-1996; 96GB-0017184.
XX
XX (DANT-) DANISCO AS.
XX
XX Gravesen TN, Rasmussen P;
XX
XX WPI; 1998-159548/14.
XX
XX N-PSDB; AAV16436.
XX
XX New Aspergillus niger beta-1,4-endo:glucanase - used for the
XX production of foodstuffs, feed and detergents and in brewing, pulp
XX and textile industries
XX
XX Claim 1; Page 57; 89pp; English.
XX
XX The endo-beta-1,4-glucanase II enzyme can be expressed in filamentous
XX fungi, plant and yeast systems. The enzyme can be used to degrade
XX glucans. This can be used for the preparation of foodstuffs, including
XX feeds, as well as in the brewing, pulp and textile industries and in the
XX formulation of detergents.
XX
XX Sequence 304 AA:
XX
XX Query Match 17.1%; Score 379.5; DB 19; Length 304;
XX Best Local Similarity 28.3%; Pred. No. 3.2e-26;
XX Matches 87; Conservative 60; Mismatches 129; Indels 31; Gaps 6;
XX
XX 4 KYLGVALPGLDIFGCDIDGSCPT-----DTSSVPLSTYKGGGAGOMKHFAEDGLNFR 58
XX 5 EMFGSNESGAEFFGNIGVWCTDIYIFPPPSAISTLI-----DKGMNFRV 49
XX
XX 59 SATMOFVLTNNVVDKLDLBNMGSYNKYVNACTLGAYCMIDMHNPARYNGSIIGCGVSD 118
XX 50 QPMERLPLDSMTGSYDEEYLANLTIVKAVTDCGALHVPNHYGRNGEIIIS---ST 105
XX
XX 119 DIFVDLWVQIAKYEDNDKIIIFGLMNEPHDIEIMWOTCOKVYTAIRKAGATSQMILLP 178
XX 106 SDFQTFMENLAGQYKNDLVAFDTNNEVHMDQDLVNLNQAALNGIRACATSOYIFYVE 165
XX
XX 179 GTFNASVETVYSTGSAELAKITNPDSSTDLIYDVHRYKIIDINNSGSHAECTDNV--DA 236
XX 166 GNSWTGAMTWADV--DNMKMLTDPDPE--DKIYVEMHQYLDSDSGSETCVSETICKER 220
XX
XX 237 FNDFAFMLROKROAIISETGASMEPSGCMTAFCAONKAISNSDVYIGFVGMGSGDPTS 296
XX 221 VTEATQWLKDKKKVGFGEYAGGSNDVCRSAVSGMLEYMANNTDVMKASWMAAGPWWGD 280
XX
XX 297 YILFTLP 303
XX 281 YIFSMEP 287
XX
XX RESULT 9
XX AAM46814
XX ID AAM46814 standard; Protein; 332 AA.
XX
XX AAM46814;

```

```

XX 18-AUG-1998 (first entry)
XX
XX Endo beta-1,4-gluconase.
XX
XX Endo-beta-1,4-glucanase II; transgenic; degradation; glucan.
XX
XX Aspergillus niger.
XX
XX Key Location/Qualifiers
XX Key Peptide 1..28
XX Protein 29..332
XX /note- "Signal peptide"
XX /note- "Mature protein"
XX
XX MO9806858-A1.
XX
XX 19-FEB-1998.
XX
XX 11-AUG-1997; 97WO-EP04415.
XX
XX 15-AUG-1996; 96GB-0017184.
XX
XX (DANT-) DANISCO AS.
XX
XX Gravesen TN, Rasmussen P;
XX
XX WPI; 1998-159548/14.
XX
XX N-PSDB; AAV16444.
XX
XX New Aspergillus niger beta-1,4-endo:glucanase - used for the
XX production of foodstuffs, feed and detergents and in brewing, pulp
XX and textile industries
XX
XX Disclosure; Page 60-62; 89pp; English.
XX
XX The endo-beta-1,4-glucanase II enzyme can be expressed in filamentous
XX fungi, plant and yeast systems. The enzyme can be used to degrade
XX glucans. This can be used for the preparation of foodstuffs, including
XX feeds, as well as in the brewing, pulp and textile industries and in the
XX formulation of detergents.
XX
XX Sequence 332 AA:
XX
XX Query Match 17.1%; Score 379.5; DB 19; Length 332;
XX Best Local Similarity 28.3%; Pred. No. 3.7e-26;
XX Matches 87; Conservative 60; Mismatches 129; Indels 31; Gaps 6;
XX
XX 4 KYLGVALPGLDIFGCDIDGSCPT-----DTSSVPLSTYKGGGAGOMKHFAEDGLNFR 58
XX 33 EMFGSNESGAEFFGNIGVWCTDIYIFPPPSAISTLI-----DKGMNFRV 77
XX
XX 59 SATMOFVLTNNVVDKLDLBNMGSYNKYVNACTLGAYCMIDMHNPARYNGSIIGCGVSD 118
XX 78 QPMERLPLDSMTGSYDEEYLANLTIVKAVTDCGALHVPNHYGRNGEIIIS---ST 133
XX
XX 119 DIFVDLWVQIAKYEDNDKIIIFGLMNEPHDIEIMWOTCOKVYTAIRKAGATSQMILLP 178
XX 134 SDFQTFMENLAGQYKNDLVAFDTNNEVHMDQDLVNLNQAALNGIRACATSOYIFYVE 193
XX
XX 179 GTFNASVETVYSTGSAELAKITNPDSSTDLIYDVHRYKIIDINNSGSHAECTDNV--DA 236
XX 194 GNSWTGAMTWADV--DNMKMLTDPDPE--DKIYVEMHQYLDSDSGSETCVSETICKER 248
XX
XX 237 FNDFAFMLROKROAIISETGASMEPSGCMTAFCAONKAISNSDVYIGFVGMGSGDPTS 296
XX 249 VTEATQWLKDKKKVGFGEYAGGSNDVCRSAVSGMLEYMANNTDVMKASWMAAGPWWGD 308
XX
XX 297 YILFTLP 303
XX 309 YIFSMEP 315
XX
XX

```

PR 30-SEP-1998; 98JP-0377864.  
 XX  
 PA (YASO/) YASOKAWA D.  
 XX  
 XX WPI: 2000-402850/35.  
 DR N-PSDB: AAA53338.  
 XX  
 PT A gene encoding endoglucanase -  
 XX  
 PS Claim 1: Page 8-9; 10pp; Japanese.  
 XX  
 CC The present sequence represents an endoglucanase amino acid sequence. The  
 CC invention relates to the gene and the protein encoded by it, having  
 CC endoglucanase activity. Also included in the invention is an expression  
 CC vector containing the gene sequence, and a fungus such as *Aspergillus*  
 CC transformed by the expression vector. Endoglucanase is used to hydrolyse  
 CC cellulose.  
 CC  
 XX  
 SQ Sequence 390 AA:  
 Query Match 28.0%; Score 620; DB 21; Length 390;  
 Best Local Similarity 44.3%; Pred. No. 3,7e+48;  
 Matches 136; Conservative 55; Mismatches 98; Indels 18; Gaps 11;  
 OY 2 KIKYL-GVAIPGIDFGCDIDGSCPTDTSVPLLSYKGGDAGOMKFAEDDGLNFRISA 60  
 DB 78 RLPLYGQVNTAGYDFVDTGTG-TGCGVPP-----PASQYAHFA-NEGANLFRIPF 127  
 OY 61 TWQVLTNTVYDGLDELNMG-SYKVVNACLEFGAYCMIDMHPARYNGSIIGOGVSDD 119  
 DB 128 AMQ-LMTPTLGSSINOTFFQSEYNPVQAALATGAVYIVDLHNVARVNCQIIGCGGPTNA 186  
 OY 120 IEPDLWQIAKKYEDNDKIFGLMNEPHLD-TEIMAQTCQKYTAIRKAGATSQMLLP 178  
 DB 187 QFASITPTQLSTYGNNEKVFGLMNEPHDLSIPENWADSLQYVNAVRAAGSTN-YLLP 245  
 OY 179 GTNFASVETVSTGSAEALCKITNPDSGLDYEDVHKYLDINNSGSHACTTDNDVAFN 238  
 DB 246 GSSWASAGQALPTFAGGYLLDITPLDGTNKLIFDVHKYLDSDNSGTHSCVNTNCGVLQ 304  
 OY 239 DFDMDTRON-KROAITSET-GASMEPSQMTAFCAQNKAIKSESDVYIGVWGAGSFTDS 296  
 DB 305 THVTWLOQNGNRQALLSETGGSSDSCETVVAQELAFVQANRNNIAGFAIMAGADTTP 364  
 OY 297 YILTLTP 303  
 DB 365 YVLSVTP 371  
 RESULT 7  
 AAEL2786  
 ID AAEL2786 standard; Protein; 335 AA.  
 XX  
 AC AAEL2786;  
 XX  
 DT 15-JAN-2002 (first entry)  
 DE Talaromyces emersonii beta-glucanase CEA protein.  
 XX  
 XX Beta-glucanase; CEA; antihypertensive; fungicide; hyperlipaemia; animal feed;  
 KW fabric; edible foodstuff; textile; brewing; distilling; bioremediation;  
 KW dental hygiene; leather-treatment; paper manufacture; plant extract;  
 KW detergent treatment; baking.  
 XX  
 OS Talaromyces emersonii.  
 XX  
 PN WO200170998-A1.  
 XX  
 PD 27-SEP-2001.  
 XX  
 PF 20-MAR-2001; 2001WO-EP03174.  
 XX  
 PR 20-MAR-2000; 2000EP-0302263.

XX  
 XX (STAM ) DSM NV.  
 PA  
 PI Van Den Hombergh JPTW, Van Der Laan J, Daran JG, Herweijer MA;  
 PI Teufel DP;  
 XX  
 DR WPI: 2001-648392/74.  
 DR N-PSDB: AAD0928.  
 XX  
 PT New polypeptide for treating hyperlipemia and/or high serum  
 PT cholesterol and triglyceride levels, comprises the beta-glucanase  
 PT protein obtainable from *Talaromyces* fungus -  
 XX  
 PS Claim 3; Page 66; 76pp; English.  
 CC  
 CC The invention relates to a beta-glucanase polypeptide obtainable from a  
 CC fungus of the genus *Talaromyces*, e.g. *T. emersonii*, having endoglucanase  
 CC activity (EC 3.2.1.4). Beta-glucanase polypeptide is useful for the  
 CC manufacture of a medicament for treating hyperlipaemia, high serum  
 CC cholesterol and triglyceride levels. Beta-glucanase polypeptide are  
 CC useful for treating fungal or plant materials (plant pulp, plant  
 CC extracts), edible foodstuffs or ingredients, or fabrics, textiles or  
 CC clothes containing plant materials. Beta-glucanase polypeptide is  
 CC useful for reducing the viscosity of a plant material, for cleaving  
 CC beta-D-glucan polymers in the plant material, and for processing plant  
 CC pulp, juice or extract, by incubating the pulp, juice or extract.  
 CC beta-glucanase polypeptide is applicable in brewing, distilling,  
 CC bioremediation, dental hygiene, leather-treatment, paper manufacture,  
 CC textile treatment or manufacture, baking or bread making, washing or  
 CC detergent treatment, treating flower bulbs or in animal feed.  
 CC beta-glucanase polypeptide is also useful during the production of milk  
 CC substituted from soybean. The present sequence is *Talaromyces emersonii*.  
 CC  
 XX  
 SQ Sequence 335 AA:  
 Query Match 17.6%; Score 389; DB 22; Length 335;  
 Best Local Similarity 30.1%; Pred. No. 5e+27;  
 Matches 89; Conservative 54; Mismatches 123; Indels 30; Gaps 7;  
 OY 10 IPGIDFGCDIDGSCPTDTSVPLLSYKGGDAGOMKFAEDDGLNFRISATQVFNNT 69  
 DB 50 IPGVE---GTDYTPP-NTSAIQILI-----DQGMNIFRVEFLMERVNPQ 90  
 OY 70 VDGKDELNMGSYKVVNACLEFGAYCMIDMHPARYNGSIIGOGVSDDIFVWLWQIA 129  
 DB 91 MTGPVDSAYFQGYQVNTYTHSGASAVIDPHNFGRTYNNIIS-----SPDDFOTFNHTIA 146  
 OY 130 KYEDNDKIFGLMNEPHDIDIEIMQTCQKYVTAIRKAGATSQMLLPSTNFASVETVY 189  
 DB 147 SNFADNDNVLEFDNNEVHDMDESLVQNLQALIDGIRAGATSQYIFVEGNSWTGAWTWT 206  
 OY 190 STGSAEALCKITNPDSGLDYEDVHKYLDINNSGSHACTTDNV--DANDRADMDLRON 247  
 DB 207 QVN--DAMANWLTDPQNK--IYEMHQYLDSDSGTSIDCVNSTIGDREVESATWMLKON 261  
 OY 248 KROAITSETGASMEPSQMTAFCAQNKAIKSESDVYIGVWGAGSFTDSYILTLTP 303  
 DB 262 GKRAILGEYAGANSVCTAVTGMIDYLANNTDVTGCAIMWAGPMMGDIIFSNRP 317  
 RESULT 8  
 AAM46807  
 ID AAM46807 standard; Protein; 304 AA.  
 XX  
 AC AAM46807;  
 XX  
 DT 18-AUG-1998 (first entry)  
 DE Endo-beta-1,4-glucanase II enzyme.  
 XX  
 XX Endo-beta-1,4-glucanase II; transgenic; degradation; glucan.  
 XX

```

Db 93 VRFAGVNIAGFDGCTTDCGCTVTSKYVPPLKNTFGSNYPGJICOMGHFVNEDEMTIFRL 152
OY 59 SATWQFVLNNTVDGKLDELNMGSYKVVNACLETGAYCMIIDMHNFPARYNGGIIQGGVSD 118
Db 153 PVGMQYLVNMMNLGCLNSTSKYDQVGLSLGAVCTVDIHNYARWNGGIIQGGGPTN 212
OY 119 DIFVDLWQIAKYYEDNDKIIIFGLMNEPHDIEIWMOTCKVYVTAIRKAGATSQMILLP 178
Db 213 AOFSLMSQASKYASQSRVWFGLMNEPHDVNIMTMATVGEVYTAIRMGATSOFTSLP 272
OY 179 GTNFSVETVYVSTGSAEALGKITPDGSTDLLYFDVHKYLDINNCSGHAECTTNDV-AF 237
Db 273 GNDWQSAQAFISDSAAALSGVTNPDSGTMLIFDVHKYLDSDNSGTHAECTTNNIDGAF 332
OY 238 NDPAWMLRONKROAIISPTGSMPEPCMTAFCAQNKAISENSDYVIGFVWGAGSFDSTY 297
Db 333 SPLATWMLRONKROAILTETGGGNVQSCIDMKCQOIQYLNQNSDVYLGIVWGAGSFDSTY 392
OY 298 ILTLTPPLGKPGNYTDNKLMEICI 320
Db 393 VLTEPTSSGNSMTDTSLSVSSL 415

```

RESULT 5  
AAM02032  
ID AAM02032 standard; Protein; 418 AA.

```

AC AAM02032:
XX 28-OCT-1996 (first entry)
XX Trichoderma endoglucanase II.
DE Trichoderma endoglucanase II.
XX Endoglucanase II: Egit; cellulase; cellulose; denim;
KM stonewashing; dye redeposition; backstaining.
XX Trichoderma longibrachiatum.
OS
XX
FH Key 1 Location/Qualifiers
FT Peptide 1..21
FT Protein /label= sig_peptide
FT /label= Mat_protein
FT Domain 22..57
FT /label= Cellulose_binding_domain
FT Region 58..91
FT /label= Linker_region
FT Domain 92..418
FT /label= Catalytic_core_domain
FT /note= "catalytic core domain is the preferred
domain for use in constructs of the
invention"

```

```

XX W09623928-A1.
XX 08-AUG-1996.
XX 29-JAN-1996; 96WO-US00977.
XX 01-FEB-1995; 95US-0382452.
XX (GEMV ) GENENCOR INT INC.
XX Clarkson KA, Collier KD, Fowler T, Larens E, Ward M;
XX WPI; 1996-371466/37.
XX N-PSDB; AAT32223.
XX Treatment of cellulose-contg. fabrics such as denim, e.g.
XX stonewashing - using truncated cellulase enzyme to increase
XX abrasion and give reduced redeposition of dye
XX

```

```

PS Disclosure; Fig 4A-4C; 124pp; English.
XX The amino acid sequences for Trichoderma longibrachiatum
CC cellulohydrolase I (CBHI) (AAM02022), CBHI (AAM02025), endoglucanase I
CC (EBI) (AAM02029), Egit (AAM02032) and EgitI (AAM02034), were deduced from
CC the respective genomic DNA sequences (AAT32220-24). The CBHI, CBHI1,
CC Egit and EgitI enzymes have catalytic core domains useful for reducing
CC dye redeposition (backstaining) on cellulose-contg. fabrics such as
CC denim, whilst maintaining or increasing abrasion during stonewashing.
CC Truncated enzymes comprising these catalytic core domains can be obtd.
CC by proteolysis of the complete enzyme or by inserting the appropriate
CC DNA fragment into a vector, using this to transform a Trichoderma
XX sp. host cell, and recovering the recombinant core domain.
SO Sequence 418 AA:

```

Query Match 42.6%; Score 942.5; DB 17; Length 418;  
Best Local Similarity 51.7%; Pred. No. 9e-78;  
Matches 167; Conservative 61; Mismatches 90; Indels 5; Gaps 2;

```

OY 3 IKYLGVAIPGIDFGCDIDGSCPTDTSVPLSYKGG----DGAGOMKHFAEDDGLNVERI 58
Db 93 VRFAGVNIAGFDGCTTDCGCTVTSKYVPPLKNTFGSNYPGJICOMGHFVNEDEMTIFRL 152
OY 59 SATWQFVLNNTVDGKLDELNMGSYKVVNACLETGAYCMIIDMHNFPARYNGGIIQGGVSD 118
Db 153 PVGMQYLVNMMNLGCLNSTSKYDQVGLSLGAVCTVDIHNYARWNGGIIQGGGPTN 212
OY 119 DIFVDLWQIAKYYEDNDKIIIFGLMNEPHDIEIWMOTCKVYVTAIRKAGATSQMILLP 178
Db 213 AOFSLMSQASKYASQSRVWFGLMNEPHDVNIMTMATVGEVYTAIRMGATSOFTSLP 272
OY 179 GTNFSVETVYVSTGSAEALGKITPDGSTDLLYFDVHKYLDINNCSGHAECTTNDV-AF 237
Db 273 GNDWQSAQAFISDSAAALSGVTNPDSGTMLIFDVHKYLDSDNSGTHAECTTNNIDGAF 332
OY 238 NDPAWMLRONKROAIISPTGSMPEPCMTAFCAQNKAISENSDYVIGFVWGAGSFDSTY 297
Db 333 SPLATWMLRONKROAILTETGGGNVQSCIDMKCQOIQYLNQNSDVYLGIVWGAGSFDSTY 392
OY 298 ILTLTPPLGKPGNYTDNKLMEICI 320
Db 393 VLTEPTSSGNSMTDTSLSVSSL 415

```

RESULT 6

AAB03663  
ID AAB03663 standard; Protein; 390 AA.

```

AC AAB03663:
XX 25-SEP-2000 (first entry)
XX Endoglucanase protein sequence.
XX Endoglucanase; cellulose hydrolysis.
XX Corticium rolfsii.
XX
FH Key 1 Location/Qualifiers
FT Peptide 1..17
FT Protein /label= "Signal peptide"
FT /label= "Endoglucanase"
FT /note= "The mature protein sequence is claimed with
the signal sequence optionally present"

```

```

XX JP2000106887-A.
XX 18-APR-2000.
XX 30-SEP-1998; 98JP-0377864.
XX

```



DR WPI: 1998-250959/22.  
 DR N-PSDB: AAV29397.  
 XX  
 PT Regulatory sequence for Trichoderma viride derived cellulase cbh1  
 PT gene - for producing Humicola insolens derived endo-glucanase  
 XX  
 PS Disclosure: Pages 48-51; 92pp; Japanese.  
 XX  
 CC The present sequence appears in the specification. The specification  
 CC describes a new regulatory sequence for Trichoderma viride derived  
 CC cellulase cbh1 gene and the establishment of a system for mass producing  
 CC cellulase in moulds such as T. viride. As the regulatory sequence of  
 CC cbh1 genes originating in T. viride can highly express objective  
 CC proteins, proteins such as cellulase can be expressed. An expression  
 CC vector containing the regulatory sequence and Humicola insolens derived  
 CC endo-glucanase NCE4 DNA was produced, and used to produce endo-glucanase  
 CC at 15 grams per litre.  
 CC  
 XX  
 SQ Sequence 418 AA:  
 Query Match 42.7%; Score 944.5; DB 19; Length 418;  
 Best Local Similarity 52.3%; Pred. No. 5,9e-78;  
 Matches 169; Conservative 57; Mismatches 92; Indels 5; Gaps 2;  
 QY 3 IKYLGVAIPGIDGCDIDGSCPTDTSVPLISYKRG---DGAGOMKHFADDDGLNFR1 58  
 DB 93 VRFAGVNIAGFDECTTDCGTCVTSKYRPLKNTFGANNYPDGIOMOHFVNDGMTIFRL 152  
 QY 59 SATMOFLNNTVDDKLDLMMGSKYNNVNCLEFGACMIDMNFARVNGIIGCGVSD 118  
 DB 153 PVGMQYLVNNNGGLTDLSTISKTDLVGCLSGVYCIIDIHVYARMNGIIGCGPTN 212  
 QY 119 DIFVDLWQIAKXYEDNDKIIIFGLMNEPHDIEIMATOCOKVYTAIRKAGATSOMLLP 178  
 DB 213 AQTSLMSQSLASKYASQSRWFGIMNEPHDVNITMTATVQEVYTAIRNAGATSOYISLP 272  
 QY 179 GTNFASVETVYVSGSABALGKITNPDSGTDLLYFDVHKYIDINNSGSHAECTTDNDV-AF 237  
 DB 273 GNDYQSAAPAFISDGAALSGVTNPDSGTNLLIFDVHKYIDSDNSGTHAECTTNNDGAF 332  
 QY 238 NDFADHLRQNRKRAIISSETGASMEPSCMTAFCAQNKAISENSDVYIFGVGAGSFPDTSY 297  
 DB 333 APLATWLRQNNRRAIILETGGGNVQSCIDLCQOIYLNONSVDYIAGAGAGSFPDTSY 392  
 QY 298 ILTLTPLGKPGNVTDNKLMNECI 320  
 DB 393 ILTETPTGSGNSWTDTSLVSSCL 415  
 RESULT 2  
 AAM97208 ID AAM97208 standard; Protein; 418 AA.  
 AC AAM97208;  
 XX  
 DT 07-MAY-1999 (first entry)  
 DE A cellulase enzyme designated SCE-3.  
 XX  
 KM Cellulase enzyme; cellulose-containing fibre; bleaching; denim-dyed;  
 KM fluff elimination; weight loss treatment; deacetylated triacetate rayon;  
 KM SCE-3.  
 XX  
 OS Trichoderma viride.  
 XX  
 FH Key Location/Qualifiers  
 FH Peptide 1..21  
 FT Protein /note="signal peptide"  
 FT Protein 22..418  
 XX /note="mature protein"  
 PN W09854332-A1.  
 XX

PD 03-DEC-1998.  
 XX  
 PF 27-MAY-1998; 98WO-JP02326.  
 XX  
 PR 27-MAY-1997; 97JP-0137258.  
 XX  
 PA (MEIJ) MEIJI SEIKA KAISHA LTD.  
 XX  
 PI Aoyagi K., Koda J., Kono T., Murakami T., Nakamura Y;  
 PI Sato Y., Sumida N., Watanabe M;  
 XX  
 DR WPI: 1999-070218/06.  
 DR N-PSDB: AAX15437.  
 XX  
 PT Cellulose preparation containing highly active cellulase SCE3  
 PT e.g. in treating cellulose-containing fibres to enable fluff  
 PT elimination, weight loss and bleaching, and in weight loss treatment  
 PT of deacetylated triacetate rayon  
 XX  
 PS Claim 1; Page 24-26; 44pp; Japanese.  
 XX  
 CC The present sequence represents a cellulase enzyme SCE-3. The  
 CC cellulase is used in the methods of the invention for treatment  
 CC of cellulose-containing fibres, for bleaching denim-dyed  
 CC cellulose-containing fibres, for eliminating fluffs from  
 CC cellulose-containing fibres, for weight loss treatment of  
 CC cellulose-containing fibres and of deacetylated triacetate  
 CC rayon, all by contacting the preparation with such fibres.  
 CC It is useful in the textile and related industries.  
 CC  
 XX  
 SQ Sequence 418 AA:  
 Query Match 42.7%; Score 944.5; DB 20; Length 418;  
 Best Local Similarity 52.3%; Pred. No. 5,9e-78;  
 Matches 169; Conservative 57; Mismatches 92; Indels 5; Gaps 2;  
 QY 3 IKYLGVAIPGIDGCDIDGSCPTDTSVPLISYKRG---DGAGOMKHFADDDGLNFR1 58  
 DB 93 VRFAGVNIAGFDECTTDCGTCVTSKYRPLKNTFGANNYPDGIOMOHFVNDGMTIFRL 152  
 QY 59 SATMOFLNNTVDDKLDLMMGSKYNNVNCLEFGACMIDMNFARVNGIIGCGVSD 118  
 DB 153 PVGMQYLVNNNGGLTDLSTISKTDLVGCLSGVYCIIDIHVYARMNGIIGCGPTN 212  
 QY 119 DIFVDLWQIAKXYEDNDKIIIFGLMNEPHDIEIMATOCOKVYTAIRKAGATSOMLLP 178  
 DB 213 AQTSLMSQSLASKYASQSRWFGIMNEPHDVNITMTATVQEVYTAIRNAGATSOYISLP 272  
 QY 179 GTNFASVETVYVSGSABALGKITNPDSGTDLLYFDVHKYIDINNSGSHAECTTDNDV-AF 237  
 DB 273 GNDYQSAAPAFISDGAALSGVTNPDSGTNLLIFDVHKYIDSDNSGTHAECTTNNDGAF 332  
 QY 238 NDFADHLRQNRKRAIISSETGASMEPSCMTAFCAQNKAISENSDVYIFGVGAGSFPDTSY 297  
 DB 333 APLATWLRQNNRRAIILETGGGNVQSCIDLCQOIYLNONSVDYIAGAGAGSFPDTSY 392  
 QY 298 ILTLTPLGKPGNVTDNKLMNECI 320  
 DB 393 ILTETPTGSGNSWTDTSLVSSCL 415  
 RESULT 3  
 AAR79540 ID AAR79540 standard; Protein; 418 AA.  
 AC AAR79540;  
 XX  
 DT 14-JAN-1996 (first entry)  
 DE Endoglucanase-II protein sequence.  
 XX  
 KM Endoglucanase-II; cellulase complex; feed-additive.  
 XX

GenCore version 5.1.4-PS.4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 9, 2003, 15:10:20 ; Search time 62.1804 Seconds  
(without alignments)  
897.904 Million cell updates/sec

Title: US-10-028-245-2

Perfect score: 2214

Sequence: 1 GKIKYGVAFPCIDPCDID.....GLTGVLFVVALGYLVAF 419

Scoring table:

BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :  
1: /SID2/gcgdata/geneeq/geneeq-emb1/AA1980.DAT:\*  
2: /SID2/gcgdata/geneeq/geneeq-emb1/AA1981.DAT:\*  
3: /SID2/gcgdata/geneeq/geneeq-emb1/AA1982.DAT:\*  
4: /SID2/gcgdata/geneeq/geneeq-emb1/AA1983.DAT:\*  
5: /SID2/gcgdata/geneeq/geneeq-emb1/AA1984.DAT:\*  
6: /SID2/gcgdata/geneeq/geneeq-emb1/AA1985.DAT:\*  
7: /SID2/gcgdata/geneeq/geneeq-emb1/AA1986.DAT:\*  
8: /SID2/gcgdata/geneeq/geneeq-emb1/AA1987.DAT:\*  
9: /SID2/gcgdata/geneeq/geneeq-emb1/AA1988.DAT:\*  
10: /SID2/gcgdata/geneeq/geneeq-emb1/AA1989.DAT:\*  
11: /SID2/gcgdata/geneeq/geneeq-emb1/AA1990.DAT:\*  
12: /SID2/gcgdata/geneeq/geneeq-emb1/AA1991.DAT:\*  
13: /SID2/gcgdata/geneeq/geneeq-emb1/AA1992.DAT:\*  
14: /SID2/gcgdata/geneeq/geneeq-emb1/AA1993.DAT:\*  
15: /SID2/gcgdata/geneeq/geneeq-emb1/AA1994.DAT:\*  
16: /SID2/gcgdata/geneeq/geneeq-emb1/AA1995.DAT:\*  
17: /SID2/gcgdata/geneeq/geneeq-emb1/AA1996.DAT:\*  
18: /SID2/gcgdata/geneeq/geneeq-emb1/AA1997.DAT:\*  
19: /SID2/gcgdata/geneeq/geneeq-emb1/AA1998.DAT:\*  
20: /SID2/gcgdata/geneeq/geneeq-emb1/AA1999.DAT:\*  
21: /SID2/gcgdata/geneeq/geneeq-emb1/AA2000.DAT:\*  
22: /SID2/gcgdata/geneeq/geneeq-emb1/AA2001.DAT:\*  
23: /SID2/gcgdata/geneeq/geneeq-emb1/AA2002.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	944.5	42.7	418	19	AAW57421
2	944.5	42.7	418	20	AAW57208
3	942.5	42.6	418	16	AAW79540
4	942.5	42.6	418	16	AAW77264
5	942.5	42.6	418	17	AAW02032
6	620	28.0	390	21	AAW03663
7	389	17.6	335	22	AAW12786
8	379.5	17.1	304	19	AAW46807
9	379.5	17.1	332	19	AAW46814
10	361	16.3	331	18	AAW14596

11	361	16.3	331	18	AAW15563	Cellulase polypept
12	184	8.3	800	15	AAW47496	Translated sequenc
13	175	7.9	471	19	AAW56742	Ornomyces cellu
14	163.5	7.4	477	21	AAW08774	Amino acid sequenc
15	162.5	7.3	582	22	AAW62467	P. rhizinflata egl
16	159	7.2	390	22	AAW62468	P. rhizinflata egl
17	159	7.2	394	22	AAW62471	P. rhizinflata egl
18	157.5	7.1	172	19	AAW46819	Endo-beta-1,4-gluc
19	151.5	6.8	452	15	AAW49102	Translated sequenc
20	151	6.8	409	9	AAW1843	Sequence of alkali
21	147	6.6	410	18	AAW12378	P300-CE1B fusion c
22	147	6.6	411	18	AAW12379	P300-CE1B fusion c
23	147	6.6	411	18	AAW12381	P300-CE1B fusion c
24	147	6.6	412	18	AAW12380	P300-CE1B fusion c
25	146.5	6.6	360	18	AAW34566	Thermotoga OCI/4V
26	146.5	6.6	360	19	AAW49870	Thermotoga OCI/4V
27	145	6.5	317	20	AAW39338	Carboxymethyl cell
28	145	6.5	317	20	AAW23764	A carboxymethyl ce
29	145	6.5	317	21	AAW56814	T. maritima thermo
30	145	6.5	317	22	AAW08543	Thermotoga maritim
31	140	6.3	551	18	AAW18790	Corrected Bacillus
32	136.5	6.2	400	18	AAW23601	Bacillus agaradher
33	136.5	6.2	400	18	AAW22521	Bacillus agaradher
34	136.5	6.2	400	19	AAW57431	Bacillus agaradher
35	132.5	6.0	462	19	AAW57433	Cloned alkaline en
36	131	5.9	467	17	AAW05731	Cellulase. Bacill
37	131	5.9	467	17	AAW00382	Bacillus cellulase
38	129	5.8	1010	19	AAW34989	Teredinibacter end
39	121	5.5	531	16	AAW01503	60 kd endoglucans
40	117	5.3	518	22	AAW06928	Micromonospora car
41	116.5	5.3	499	14	AAW42122	NK-1 cellulase. B
42	116.5	5.3	1063	23	AAW54168	Lactococcus lactis
43	114.5	5.2	472	12	AAW37243	Globodera rostochi
44	114	5.1	532	19	AAW13329	Endoglucanase enco
45	113.5	5.1	406	11	AAW08199	Neutral cellulase

#### ALIGNMENTS

RESULT 1	AAW57421	standard; Protein; 418 AA.
ID	AAW57421	
XX	AAW57421	
AC	AAW57421	
XX		
DT	02-SEP-1998	(first entry)
XX		
DE		Amino acid sequence of the specification.
XX		
KW		Regulatory sequence; cellulase cbh1 gene; mass production;
KW		Humicola insolens; endo-glucanase NCE4.
XX		
OS		Trichoderma viride.
XX		
FH	Key	Location/Qualifiers
FT	Peptide	1..21
FT		/note="signal peptide"
FT	Protein	22..418
FT		/note="mature protein"
XX		
PN	WO9811239-A1	
XX		
PD	19-MAR-1998.	
XX		
PF	16-SEP-1997;	97WO-IP03268.
XX		
PR	13-SEP-1996;	96JP-0243695.
XX		
PA	(MEIT ) MEITI SEIKA KAISHA LTD.	
XX		
PI	Toyagi K, Moriya T, Murakami T, Sumida N, Watanabe M;	

**THIS PAGE BLANK (USPTO)**

KEYWORDS EST.  
SOURCE African malaria mosquito.  
ORGANISM Anopheles gambiae

Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
Neoptera; Endopterygota; Diptera; Nematocera; Culicoidae;  
Anopheles.

REFERENCE 1 (bases 1 to 553)

AUTHORS Holt, R.A., Lin, J.-J., Murphy, S.D., Evans, C.A., Kraft, C.L., Charlab  
R., Collins, F.H., Venter, J.C. and Hoffman, S.L.

TITLE Celera Anopheles gambiae EST project  
JOURNAL Unpublished (2002)

COMMENT

Contact: Holt R.A.  
Celera Genomics

45 W. Gude Dr., Rockville, MD 20850, USA

Tel: 2404533151

Fax: 2404534580

Email: HOLTRA@celera.com

Plate: NU010418A row: D column: 13

Seq primer: MJ3 Reverse.  
Location/Qualifiers

1. 553  
/organism="Anopheles gambiae"  
/strain="RSP-St (Reduced susc. to Permethrin - std.  
chromosome)"  
/db\_xref="taxon:7165"  
/clone="19600449666079"  
/clone.lib="A.Gam.ad.cdna1"  
/dev\_stage="Adult"  
/lab\_host="DH10b"

/note="Vector: pSport1; Site\_1: SalI; Site\_2: NotI; Whole  
adult mosquitoes (mixed sex) frozen on liquid nitrogen.  
cDNA inserts >500 bp cloned directionally into pSport 1.  
Not 1 site is 3'. Clones available through the Malaria  
Research and Reference Reagent Resource Center  
(www.malaria.mr4.org)."

BASE COUNT 107 a 170 c 168 g 108 t  
ORIGIN

Query Match 3.0%; Score 55; DB 13; Length 553;  
Best Local Similarity 48.0%; Pred. No. 0.28;

Matches 157; Conservative 0; Mismatches 170; Indels 0; Gaps 0;

```

QY 700 GGTCTGTAACGCTGTCTGAGACGGGCGCTACTGATGATGACATGCACAATTTC 759
      |||||  |||  |||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 181 GGTGCGCATCAACGATCCGTTGATCGCGCTGACATGATGATGATGATGATGATGAT 240
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
QY 760 CCGCTACACGGGCGCATGCGGCCAGGAGCGCTGTGCGACATCTTTGTGACCT 819
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 241 CTCGACGACGGCTGCTTCAAGGGCGAGGTGTCCGCCAGGCGCTGCTGTGTGAA 300
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
QY 820 CTGGGTCCAGATCGCAAGTACTAGAGACAGACAGATCATCTTTGGCCTGATGAA 879
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 301 CCGCCAGAAAGATTCGCTGTTCAGAGAGCGGCGAGCCGAAAGCGATCCGTGGGTAAGC 360
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
QY 880 CGAGCCGACAGACTGACATGAGATCTGGGGGACAGCGTGCAGAAAGTCTGCTACATGC 939
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 361 CCGCGCGGAGTACGCTGTGAGTGCAGCGGTGTTCACACACGACGAGAGGATCCGCGC 420
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
QY 940 GATCCGAAAGGCGGCGGCGACCTGCGAGATGATCTCTGCGGAGAACCACTTTGCCAG 999
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 421 CCATCTGGAGGGTGGCGGCAAGAGGTCAATCTGCGCCCTCCGCGCAATGCGCGCA 480
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
QY 1000 CGTCGAGAGTATGTGTCACACTGGCAG 1026
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 481 TGTTCGCTGCTGGTCTCAACCTGGAG 507
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||

```

Search completed: May 15, 2003, 19:41:57  
Job time : 2417.97 secs

RESULT 13					
LOCUS	AG085797	720 bp	DNA	linear	GSS 03-NOV-2001
DEFINITION	Pan troglodytes DNA, clone: PTB-083P07.F, genomic survey sequence.				
ACCESSION	AG085797				
VERSION	AG085797.1	GI:16637599			
KEYWORDS	GSS.				
SOURCE	Pan troglodytes male lymphoblast DNA, clone_lib:PTB Chimpanzee Male Pan library clone:PTB-083P07.F. Pan troglodytes				
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homidae; Pan.				
REFERENCE	Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T., Totoki,Y., Matsubae,H. and Sakaki,Y. BAC end sequences of library PTB Unpublished 2 (bases 1 to 720) Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T., Totoki,Y., Matsubae,H. and Sakaki,Y. Direct Submission Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical and Chemical Research (RIKEN), Genomic Sciences Center (GSC), 1-7-22 Shohitcho-Tsurumi-Ku, Yokohama, Kanagawa 230-0045, Japan (E-mail:himpbes@gsc.riken.go.jp, url:http://bgp.gsc.riken.go.jp/ Tel:81-45-503-9111, Fax:81-45-503-9170)				
TITLE	JOURNAL				
AUTHORS	COMMENT				
	Claones are derived from the chimpanzee BAC library PTB This BAC end was generated during the R&D process and may have higher chance of clone tracking errors. PRIMERS				
	Sequencing: -21MJ3				
FEATURES	LIBRARY				
	Vector : pKS145				
	R.Site 1 : SacI				
	R.Site 2 : SacI.				
	Location/Qualifiers				
	1..720				
	/organism="Pan troglodytes"				
	/db_xref="taxon:9598"				
	/clone="PTB-083P07.F"				
	/sex="male"				
	/cell_type="lymphoblast"				
	/clone_lib="PTB Chimpanzee Male BAC library"				
	BASE COUNT	50 a	27 c	104 g	285 t
	ORIGIN				
	Query Match	3.1%; Score 55.8;	DB 17:	Length 720;	
	Best Local Similarity	46.6%;	Pred. No. 0.17;		
	Matches 174;	Conservative 0;	Mismatches: 199;	Indels 0;	Gaps 0;
OY	16 CGTTCATCCTCCGCCCTCGCTCGCTGCCTCCCTCCCTCCCTCCCATCACGGTCA	75			
Db	280 CCTCCCCTTTTTTTTCCCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCT	339			
OY	76 CCACCTCCGTATTGCCGCTCTCTGGGAGAGCATGACGACATTAACACACTGTT	135			
Db	340 CCGCTCTCTCTTCCGCCCTCCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT	399			
OY	136 TCTGTACTCTCGCTGCTGGGAGCTGCTGCTGGCATTCGCTTAGACATTTGTCG	195			
Db	400 TTTTTCCTTGCTTTCTCTCCCTTTTCTTCTTCTTCTTCTCTCTCTCTCTCTCT	459			
OY	196 TTCCTGCTTCTCTTAAATCCGTCATCTTCGATTCGATTCGATTCGATTCGAT	255			
Db	460 CTCCTCCCT	519			
OY	256 AGTGCTAATAACGACAGCTTTTCTTCCCTGCTGAACAGCTGACGATTAATAC	315			
Db	520 CCCCT	579			
OY	316 AACATCGTGAATAGGATATACCCTGGGCTCTCTCTGCTGCTGCTGCTGCTG	375			
Db	580 TTTCCTCTTCTCTCTCTCTTATATCCCTCTTATATCCCTCTTATATCCCTCT	639			

RESULT 14	CNS0166G	1201 bp	DNA	linear	GSS 26-JUL-1999
LOCUS	CNS0166G/c				
DEFINITION	Drosophila melanogaster genome survey sequence T7 end of BAC BACN15112 of DrosBAC library from Drosophila melanogaster (fruit fly), genomic survey sequence.				
ACCESSION	AL106354				
VERSION	ALI06354.1	GI:5621495			
KEYWORDS	GSS.				
SOURCE	Drosophila melanogaster.				
ORGANISM	Drosophila melanogaster. Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila. 1 (bases 1 to 1201)				
REFERENCE	Genoscope.				
AUTHORS	Direct Submission				
TITLE	Submitted (23-JUL-1999) Genoscope - Centre National de Sequençage				
JOURNAL	Bp 191 91006 EVRY Cedex - FRANCE (E-mail : segre@genoscope.cns.fr - Web : www.genoscope.cns.fr) Determination of this BAC-end sequence was carried out as part of a collaboration with the European Drosophila Genome Project (EDGP) - http://www.edgp.ebi.ac.uk -. This Drosophila melanogaster BAC library (Dros BAC) was made by Alain Billaut at CEPH (Centre d'Etude du Polymorphisme Humain) with funding provided by a MRC project grant. The DNA was prepared from embryos by Alain Bucheton and Genevieve Payan. It has been constructed in the vector pbeloBAC11.				
FEATURES	location/qualifiers				
Source	1..1201 /organism="Drosophila melanogaster" /db_xref="taxon:7227" /clone="BACN15112" /clone_1lb="DrosBAC" /plasmid="pbeloBAC11" /note="end : T7"				
BASE COUNT	350 a 219 c 250 g 236 t	146 others			
ORIGIN					
Query Match	3.1%; Score 55.8; DB 17; Length 1201;				
Best Local Similarity	38.3%; Pred. No.0.14;				
Matches	90; Conservative 47; Mismatches 99; Indels 0; Gaps 0;				
OY	2 TCGAGCCAGCGCTGCATTCCTCCTCCCTCCTCCTCCTCCTCCTCCCTTCTC 61				
DB	1165 TCYYCTGCCCTTTTCTCHCCGCCCYTYTCYCTCCCTCCTCCTCCGCCCTC 1106				
OY	62 CCCCATCACCGTCACCACTCTCTCATTCGCCCTCTCTGTGGAGCATTGACAGATT 121				
DB	1105 ACCBCCCACCTCCYCCYCCYCTTCTTCCTCCCTCCCTCCCTCCCTCCCTCC 1046				
OY	122 AACACACACTGCTTCTCTGTAATCTCGGCTGCGTGCGGCTGCTGCTGGAATCTGCTT 181				
DB	1045 SACCCTCAYGCGSKTTTTTTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT 986				
OY	182 AGCATTGTTGTTGCTTCTGTTCTCTCTTAAATCCGTCATCTTCGACAATGCT 236				
DB	985 KSBTTTARBSCTTTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 931				
RESULT 15	BM639114	553 bp	mRNA	linear	ESF 26-FEB-2002
LOCUS	17000667567887 A.Gam.ad.cdNAL Anopheles gambiae cDNA clone				
DEFINITION	19600449666079 5', mRNA sequence.				
ACCESSION	BM639114				
VERSION	BM639114.1	GI:18938625			



[illegible]

RESULT	9
LOCUS	AO788019/c
DEFINITION	AO788019 1260 bp DNA linear GSS 03-AUG-1999
ACCESSION	HS_3100_A2-G08-T7C CIT Approved Human Genomic Sperm Library D Homo
VERSION	AO788019
KEYWORDS	sapiens genomic clone Plate=3100 Col=16 Row=M, DNA sequence.
SOURCE	AO788019.1 GI:5695643
ORGANISM	GSS.
	human.
	Homo sapiens

REFERENCE	1 (bases 1 to 1260)
AUTHORS	Mahairas, G.G., Wallace, J.C., Smith, K., Swartzell, S., Holzman, T., Keller, A., Shaker, R., Furlong, J.J., Young, J., Zhou, S., Adams, M.D. and Hood, L.
TITLE	Sequence-tagged connectors: A sequence approach to mapping and scanning the human genome
JOURNAL	Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9739-9744 (1999)
MEDLINE	99380589
COMMENT	Contact: Mahairas GG, Wallace JC, Hood L

401 Queen Anne Avenue North, Seattle, WA 98109, USA  
Tel.: (206) 616-3618  
Fax: (206) 616-3887  
Email: [jwallaceu.washington.edu](mailto:jwallaceu.washington.edu)  
Clones may be purchased from Research Genetics ([info@resgen.com](mailto:info@resgen.com)).  
BAC end Web Server: <http://www.hnsc.washington.edu>  
Plate: 3100 row: M column: 16  
Seq primer: 77  
Class: BAC ends  
High quality sequence. stop: 1260. . .

```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone-"Plate=3100 Col=16 Row=M"
/clone_lib="CIT Approved Human Genomic Sperm Library D"
/sex="male"
/note="Organ: sperm; Vector: pBel0BAC11; BAC Clones in
E-coli DH10B"
BASE COUNT      510 a      569 g      36 t      92 others
ORIGIN

```

```

Query Match      3.1%; Score 56; DB 17; Length 1260;
Best Local Similarity 52.4%; Pred. No. 0.13;
Matches 122; Conservative 0; Mismatches 111; Indels 0; Gaps 0;

QY 7 CCACGCGTCGGTCAATTCCTTCCTCCGCGTCGTCGTCCTGCTCCGCTTCGCCCA 66
   || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 955 CCGGCCGCTCTCTCTCTCTCTCTCTCCGCTGCGCTCTCTCTCTCTCTCTCTCC 896
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 67 TCACCGTCACCACTCTCCCTCAATGCGGCTCTCTGCGAGCGCATGACGACCATTAACAC 126
   || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 895 TCTCTTCCCGCTCCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTC 836
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

	127	187	775
-QY	ACACGCGTTTCGTTACACCTCCGTCGCGCGCGTCGCGTGGATTCCTAGACG		
Db	CTTTCGTCGCTCTTCTCTCTCTCCGCCCTTTCTTTTCTCTGNCNCGCTTCTTC		
QY	TTTGTTGCTGTCGCTCTCTTTAATCCGACATCTCGAATCGTGGC	239	
Db	TCCTCTGCGCTCTCTTCTCTCTTCGCCCTCTGCGCTCTCTTCTCTCCGCC	723	

RESULT 10			
BG58164			
LOCUS	BG58164	652 bp	mRNA linear
DEFINITION	10240505B08.y1 C. reinhardtii CC-1690, normalized, Lambda Zap II		EST 29-MAY-2001
ACCESSION	BG58164		
VERSION	BG58164.1	GI:14239348	
KEYWORDS	EST.		
SOURCE	<i>Chlamydomonas reinhardtii</i> .		
ORGANISM	<i>Chlamydomonas reinhardtii</i>		

REFERENCE	1 (bases 1 to 652)
AUTHORS	Grossman, A., Davies, J., Federspiel, N., Harris, E., Lefebvre, P., Mcermott, J. P., Silflow, C., Stern, D. and Surzycki, R.
TITLE	Analyses of the Chlamydomonas reinhardtii genome: A model, Unicellular System for Analyzing Gene Function and Regulation in Vascular Plants: project phase 2
JOURNAL	Unpublished (2000)
COMMENT	Contact: Charles Hauser

**FEATURES**  
**SOURCE**

Duke University  
 Durham, NC 27708-1000  
 Tel.: 919 613 8159  
 Fax: 919 613 8177  
 Email: [chauser@duke.edu](mailto:chauser@duke.edu),  
[Location/Qualifiers](mailto:Location/Qualifiers)  
 1..652

/organism="Chlamydomonas reinhardtii"  
 /strain="CC-1690 wild type mt+ 21gr"  
 /db\_xref="taxon:3055"  
 /clone\_lib="C. reinhardtii CC-1690, normalized, lambda zap  
 II"  
 /note="Vector: pluescript II SK-. Site.1: EcoRI. Site.2:  
 XhoI. This library, constructed by John Davies and Jeffrey  
 McDermott, combines cDNAs from CC-1690 cells grown to  
 mid-log phase in TAP (acetate-containing) medium in the  
 light, TAP medium in the dark, HS (minimal) medium in  
 ambient levels of CO<sub>2</sub> and HS medium bubbled with 5% CO<sub>2</sub>.  
 POLYA mRNA was purified from each sample, pooled and cDNA  
 synthesized. The cDNA was directionally cloned into lambda  
 ZAP II (Stratagene) in the EcoRI (5') and XhoI (3') sites.  
 Pluescript II SK- plasmids were excised from the lambda  
 ZAP clones by superinfection with Exsist (Stratagene)  
 phage. The library was normalized using method 4 described  
 in Bonaldo et al (1996) Genome Research 6: 791-806."

	Query Match	3.1%	Score 55.8	DB 12	Length 652
	Best Local Similarity	52.3%	Pred. No. 0.18		
	Matches 123	Conservative 0	Mismatches 112	Indels 0	Gaps 0
QY	649	CAACACGGTGGACGGCAAGCTGAGACGACTCACTGCGGGCTCTTACAAACAAGTGTCTCAA	708		
Db	300	CAACACGCTGTGCGGCAAGAGATCGTGTGTACGGCTTTGCCTTTCAAGAGGACACGGG	359		
QY	709	CGCCTGTCTCGAGACGGGCGCCCTACTGATGATTGACATGCAACAATTGGCCCGCTACAA	768		
Db	360	CGACACTGCGCAGACGGCCCGCCATTGACGTGTGCAGAGGGCTCATCCGCGACGGCGCAA	419		
QY	769	CGGGGGCATTCATCGGCGCCAGGAGAGCGCTGTGCGAGACACTTTTGTGCAACTCTGGGTCCA	828		
Db	420	GTCGTGCATTTTCACACCCCGAGGTCAAGCGCCAGACAGATCTTTCGCGACCTGTCCGGCCC	479		

[illegible]

RESULT 7	CNS00418	LOCUS	DEFINITION	ACCESSION	VERSION	KEYWORDS	SOURCE	ORGANISM	REFERENCE	AUTHORS	TITLE	JOURNAL
	CNS00418	987 bp	DNA	linear	GSS	03-JUN-1999						
	Drosophila melanogaster genome survey sequence TE3 end of BAC #											
	BAR03C16 of RpCl-36 library from Drosophila melanogaster (fruit											
	fly), genomic survey sequence.											
	AL066537											
	AL066537.1	GI:4942778										
	GSS.											
	Drosophila melanogaster.											
	Drosophila melanogaster											
	Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;											
	Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;											
	Ephydroidea; Drosophilidae; Drosophila;											
	1 (bases 1 to 987)											
	Genoscope.											
	Direct Submission											
	Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :											

**COMMENT**

Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see [http://www.fruitfly.org/TheBDGP/Drosophila\\_melanogasterBAClibrary](http://www.fruitfly.org/TheBDGP/Drosophila_melanogasterBAClibrary) was prepared by Kazuhiro Oseegawa and Aaron Mammossier in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPC1-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain y2; cn bw sp, the same strain used for the BDGP's P1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at [http://bacpac.med.buffalo.edu/drosophila\\_bac.htm](http://bacpac.med.buffalo.edu/drosophila_bac.htm).

```

/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
;
/clone="BACR08C16"
/clone_1lp="RPC1-98"
/note_end : TET3
238 a 162 c 17 g 177 t 393 others
BASE COUNT
ORIGIN

```

**Query Match** 3.7%; Score 67.2; DB 1.7; Length 987;  
Best Local Similarity 23.6%; Pred. No. 00069;  
**Matches** 96; Conservative 131; Mismatches 179; Indels 0; Gaps 0;

[illegible]

RESULT 8	
CNS0181N/C	
LOCUS	1101 bp DNA linear GSS 26-JUL-1999
DEFINITION	Drosophila melanogaster genome survey sequence SP6 end of BAC BACN37P10 of DrosBAC library from Drosophila melanogaster (fruit fly) genomic survey sequence.
ACCESSION	AL108173
VERSION	AL108173.1 GI:5629077
KEYWORDS	GSS.
SOURCE	Drosophila melanogaster.
ORGANISM	Drosophila melanogaster

REFERENCE AUTHORS TITLE JOURNAL	COMMENT
1 (bases 1 to 1101) Genoscope. Submitted (23-JUL-1999) BP 191 91006, EVRY cedex - FRANCE (E-mail : segr@efgenoscope.cns.fr)	Genoscope - Centre National de Sequencage : Direct Submission - web : <a href="http://www.genoscope.cns.fr">www.genoscope.cns.fr</a> Determination of this BAC-end sequence was carried out as part of a collaboration with the European Drosophila Genome Project (EDGP) - <a href="http://www.edgp.ebi.ac.uk">http://www.edgp.ebi.ac.uk</a> . This Drosophila melanogaster BAC library (Dros BAC) was made by Alain Billaut at CEPH (Centre d'Etude du Polymorphisme Humain) with funding provided by a MRC project grant. The DNA was prepared from embryos by Alain Bucheton and Genevieve Payan. It has been constructed in the vector paeloBAC11.

FEATURES	SOURCE	Location/Qualifiers
		1. .1101
		/organism="Drosophila melanogaster"
		/db_xref="taxon:7227"
		/clone="BACN37P10"
		/clone_11b="DrosBAC"
		/plasmid="pBelobAC11"
		/note="end : Sp6"
BASE COUNT		125 a 364 c 197 g 116 t 299 others
ORIGIN		

[illegible]

QY 74 CACCACCTCTCATTCGCGCTCTCTGCGACCATGACGACCATTTACACACTG 133  
 Db 609 CACTGCTTCTACCTGCGCTCTCTGCGACCATGACGACCATTTACACACTG 668  
 QY 134 TTCTGTTACTCTGCGCTCTCTGCGACCATGACGACCATTTAGACTTTGTT 193  
 Db 669 TCT 728  
 QY 194 CGTCT 253  
 Db 729 TCT 788  
 QY 254 GTACTGTTATATACGACGACCTTTTCTCTCTCTCTCTCTCTCTCTCTCTCT 313  
 Db 789 TCT 848  
 QY 314 ACACCATCTGTTATGAGATATACCTCTCTCTCTCTCTCTCTCTCTCTCTCT 373  
 Db 849 CT 908  
 QY 374 TCGTTCT 411  
 Db 909 TCT 946  
 RESULT 5  
 AM282705  
 LOCUS 425 bp mRNA linear EST 19-JUL-2000  
 DEFINITION LG1 292.H06.g1\_A002 Light grown 1 (LG1) sorghum bicolor cDNA, mRNA  
 ACCESSION AM282705  
 VERSION AM282705.1 GI:6672549  
 KEYWORDS EST.  
 SOURCE sorghum.  
 ORGANISM sorghum bicolor  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC  
 1 (bases 1 to 425)  
 AUTHOR Cordonnier-Pratt, M.-M., Gingle, A., Marsala, C. and Pratt, L.H.  
 JOURNAL An EST database from Sorghum: light-grown seedlings  
 COMMENT Unpublished (2000)  
 CONTACT: Cordonnier-Pratt MM  
 Laboratory for Genomics and Bioinformatics  
 The University of Georgia, Department of Plant Biology  
 Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA  
 Tel: 706 542 1860  
 Fax: 706 583 0210  
 Email: mmpratt@uga.edu  
 Sequences have been trimmed to exclude PolyA, vector and regions  
 below phred quality 16. The threshold for highest quality sequence  
 is 20.  
 Seq primer: T7  
 High quality sequence start: 85  
 High quality sequence stop: 425  
 PolyA-tes:  
 FEATURES  
 source location/Qualifiers  
 1..425  
 /organism="Sorghum bicolor"  
 /db\_xref="taxon:4558"  
 /clone\_id="Light Grown 1 (LG1)"  
 /note="Organ: 10- to 14-day-old light-grown (greenhouse)  
 seedlings; Vector: Lambda Zap; Site\_1: XhoI; Site\_2: EcoRI  
 ; The library was made from poly-A RNA in the cloning  
 vector lambda zap II. Clones to be sequenced were  
 prepared by mass excision."  
 BASE COUNT 111 a 101 c 79 g 134 t  
 ORIGIN  
 Query Match 4.0%; Score 72.4; DB 10; Length 425;  
 Best Local Similarity 55.0%; Pred. No. 8e-05;  
 Matches 142; Conservative 0; Mismatches 116; Indels 0; Gaps 0;

QY 1151 GAGCCCTTCAACGACTTCGCGAGCTGAGGAGAGCAAGCGCCAGCCATCATCTCC 1210  
 Db 14 GATGCTTCTTACCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCT 73  
 QY 1211 GAACGGGGGGGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCT 1270  
 Db 74 GAGACAGGGGGGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCT 133  
 QY 1271 ATTGCGGAAAACAGGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCT 1330  
 Db 134 ATAAACCAAACTCGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCT 193  
 QY 1331 ACCTGAGCATCTTCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCT 1390  
 Db 194 AGCACTATATATCTTACAGAGACCCCAACAGAGTGGGAACTGCTGAGCTGAGCTGAGCT 253  
 QY 1391 CTGATGACAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCT 1408  
 Db 254 CTGATGACAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCT 271  
 RESULT 6  
 CNS017RP/c 1101 bp DNA linear GSS 26-JUL-1999  
 LOCUS Drosophila melanogaster genome survey sequence SP6 end of BAC  
 DEFINITION BACN37J10 of DrosBAC library from Drosophila melanogaster (fruit  
 fly), genomic survey sequence.  
 AL108415  
 ACCESSION AL108415.1 GI:5628719  
 VERSION GSS.  
 KEYWORDS Drosophila melanogaster.  
 SOURCE Drosophila melanogaster.  
 ORGANISM Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
 Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 Ephydroidea; Drosophilidae; Drosophila.  
 1 (bases 1 to 1101)  
 AUTHOR Genoscope.  
 JOURNAL Direct Submission  
 Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage;  
 BP 191 91006 Evry cedex - FRANCE (E-mail: seqref@genoscope.cns.fr)  
 - Web: www.genoscope.cns.fr  
 COMMENT Determination of this BAC end sequence was carried out as part of a  
 collaboration with the European Drosophila Genome Project (EDGP) -  
 http://www.edgp.ebi.ac.uk. This Drosophila melanogaster BAC  
 library (Dros BAC) was made by Alain Billaud at CEPH (Centre  
 d'Etude du Polymorphisme Humain) with funding provided by a MRC  
 project grant. The DNA was prepared from embryos by Alain Bucheton  
 and Genevieve Payan. It has been constructed in the vector  
 pBeloBAC11.  
 FEATURES  
 source location/Qualifiers  
 1..1101  
 /organism="Drosophila melanogaster"  
 /db\_xref="taxon:7227"  
 /clone\_id="BACN37J10"  
 /clone\_id="DrosBAC"  
 /plasmid="pBeloBAC11"  
 /note="end: SP6"  
 BASE COUNT 258 a 174 c 277 g 120 t 272 others  
 ORIGIN  
 Query Match 3.9%; Score 70.4; DB 17; Length 1101;  
 Best Local Similarity 27.6%; Pred. No. 0.00014;  
 Matches 113; Conservative 113; Mismatches 184; Indels 0; Gaps 0;  
 QY 2 TCGACCGACGGGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCT 61  
 Db 1097 TCT 1038  
 QY 62 CCCCATCAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCT 121  
 Db 1037 TCT 978  
 QY 122 AACACACACTGTTCTGTTACTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 181

```

0Y 806 ATCTTTGTCGACCTTGGGCTGCAGATGTCCTCAAGTCTCAAGAGGACACAGACAAAGATTCATC 865
Db 434 CAGTTCCGCCGGCCTTGGACGACGACTGGCCACCAAGTACCGGGCCACGAAAGAGTGCCTC 493
0Y 866 TTTGGCCCTGATGAACGAGCCGCCACGACCTCGACATTTGATCTCTGGGCCGACAGCTGCCAA 925
Db 494 TTTTGAGCTCATGAAACGAGCCGACAGACCTCGACATGAGGCTTCTGGGCGACAGACCTGCCAA 553
0Y 926 AAGCTCTGCTACTGCGATCCGAAAGGCCGGCCGACCTCGCAGATGATATCTCTGCCCCGA 985
Db 554 CAGGCTGCTACCGGCTATCCGCAATGCGCGGCCGACCAAGACGAGATGATCTCTGCGCGAGC 613
0Y 986 ACCAAGCTTGGCAGCGCTGAGA 1007
Db 614 AGCAACTTCACACAGCGCTGCGA 635

RESULT 3
BM865676 317 bp mRNA linear EST 07-MAR-2002
LOCUS mgc8003xc20f.b Magnaporthe grisea CS Uni-Zap XR library Magnaporthe
ACCESSION BM865676
VERSION BM865676.1 GI:19233358
KEYWORDS EST.
SOURCE Magnaporthe grisea.
ORGANISM Magnaporthe grisea
REFERENCE Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
AUTHORS Sordariomycetes Incertae sedis; Magnaportheaceae; Magnaporthe.
TITLE 1 (bases 1 to 317)
Ebolle,D.J., Yuan,J., Thomas,T.L., Bobrowicz,P. and Dean,R.A.
Expressed sequence tags from the rice blast fungus, Magnaporthe
grisea
JOURNAL Unpublished (2002)
COMMENT Contact: Ebbole DJ
Department of Plant Pathology & Microbiology
Texas A&M University
Peterson Bldg, MS2132, College Station, TX 77843-2132, USA
Tel.: 979 845 4831
Fax: 979 845 6483
Email: d-ebolle@tamu.edu
Chromatogram file of this sequence is available, see contact person
:Best nr hit (Nov. 11, 2001) dbj1BAA36216.1 (AB021657)
endoglucanase II [trichoderma viride] j31 3e-30
PCR primers
FORWARD: T3 primer
BACKWARD: T7 primer
Plate: mgc8003 row: C column: 20
Seq primer: T3.
FEATURES
Source Location/Qualifiers
1..317
/organism="Magnaporthe grisea"
/strain="Guy11"
/db_xref="taxon:148305"
/cdone="mgc8003xc20"
/clone_id="Magnaporthe grisea CS Uni-Zap XR library"
/sex="Mat1-2 hermaphrodite"
/cell_type="conidia"
/note="Vector: pBluescriptSK-+ Site.1: EcoRI; Site.2: XhoI
; Unidirectional cloning. EcoRI side has T3 primer and
predominantly 5' reads. T7 primer on XhoI side of insert.
Conidia library. Point inoculation of Guy11 at center of
natural agar plate. Conidia were harvested after two weeks
of growth. Sequences were processed by phredPhrap 991019
and trimmed according to pld files and for vector seqs."
BASE COUNT 70 a 84 c 108 g 55 t
ORIGIN
Query Match 6.08; Score 109.8; DB 14; Length 317;
Best Local Similarity 6.28; Pred. No. 1.7e-12;
Matches 177; Conservative 0; Mismatches 112; Indels 0; Gaps 0;
556 AGGAGGAGATGAGCCGCCGACATGACATTTGGCCGAGAGACAGCGCCCTCAACGCTTT 615

```

[illegible]

FORWARD: Universal M13 forward primer  
 BACKWARD: Universal M13 reverse primer  
 Plate: 11 row: A column: 10  
 Seq primer: M13 reverse primer  
 High quality sequence stop: 658  
 POLY-A-NO.

# FEATURES

source

Location/Qualifiers

1..658  
 /organism="Hypocrea jecorina"  
 /strain="QM9414 (ATCC26921)"  
 /db\_xref="taxon:51453"  
 /clone="tr-A0970"  
 /clone\_1lb="trEST-A"  
 /sex="Asexual"  
 /tissue\_type="Mycelia"  
 /dev\_stage="18 hr glycerol-grown culture"  
 /lab\_host="E. coli SOLR cells (kanamycin resistant)"  
 /note="Vector: pBluescript SK(+). Site\_1: EcoRI; Site\_2: XhoI; anamorph-trichoderma reesei; Cloned unidirectionally, 5' end of the cDNA cloned into EcoRI site of pBluescript. Primer: Oligo (dT). Average insert size: 1.2 kb; Uni-ZAP XR Vector system -5' adaptor sequence: 5'-GAATTCGACGAG3'-3' adaptor sequence: 5'-CTCGAGTTTCTTTTCTTTT3'."

BASE COUNT 159 a 182 c 170 g 147 t

ORIGIN

## Query Match

Best Local Similarity 28.6%; Score 521.4; DB 13; Length 658;  
 Pred. No. 2,1e-97;

Matches 558; Conservative 0; Mismatches 61; Indels 0; Gaps 0;

664 CAAGCTGAGAGCTCAACTGCGGCTCTTACACAGAGTGTCAAGCGCTGTGAGAC 723  
 1 CAAGCTGAGAGCTCAACTGCGGCTCTTACACAGAGTGTGCTGAGAC 60  
 724 GGGCGCTACTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 783  
 61 GGGCGCTACTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 120  
 784 CAAGGAGGCGTGTGAGACGATCTTGTGACATCTGCGTGTGAGATGCAAGTACTA 843  
 121 CCAGGAGGCGTGTGAGACGATCTTGTGACATCTGCGTGTGAGATGCAAGTACTA 180  
 844 CGAGGACAGACAGATCATCTTGTGATGATGATGATGATGATGATGATGATGATG 903  
 181 CGAGGACAGACAGATCATCTTGTGATGATGATGATGATGATGATGATGATGATG 240  
 904 GATCTGGCGGACGCGTGTGAGACGATCTTGTGACATCTGCGTGTGAGATGCAAGT 963  
 241 GATCTGGCGGACGCGTGTGAGACGATCTTGTGACATCTGCGTGTGAGATGCAAGT 300  
 964 GAGATGATCTCTGCGGACGATCTTGTGACATCTGCGTGTGAGATGCAAGT 1023  
 301 GAGATGATCTCTGCGGACGATCTTGTGACATCTGCGTGTGAGATGCAAGT 360  
 1024 CAGCGGGAAGCCCTGCGCAAGATTTAGCAACCCGATGGAAGCAACGATTTGTCTACT 1083  
 361 CAGCGGGAAGCCCTGCGCAAGATTTAGCAACCCGATGGAAGCAACGATTTGTCTACT 420  
 1084 TGATGTCACAGATTTGATGATGATGATGATGATGATGATGATGATGATGATGATG 1143  
 421 TGATGTCACAGATTTGATGATGATGATGATGATGATGATGATGATGATGATGATG 1203  
 1144 CAAGCTGAGAGCTCAACTGCGGCTCTTACACAGAGTGTGAGAC 1203  
 481 CAAGCTGAGAGCTCAACTGCGGCTCTTACACAGAGTGTGAGAC 540  
 1204 CATCTCCGAAAGGCGCGTGTGAGACCTTGTGATGATGATGATGATGATGATGATG 1263  
 541 TTTTGTGAAAGGCGCGCGTGTGAGACCTTGTGATGATGATGATGATGATGATGATG 600  
 1264 CAAGGCAATTAAGGAAAC 1282

DB 601 TGGCGCTAATTTAAACAC 619

## RESULT 2

LOCUS BG810475

DEFINITION mgct006xe06f Magnaporthe grisea Appressorium Stage cDNA Magnaporthe

ACCESSION BG810475

VERSION BG810475.1 GI:14181455

KEYWORDS EST.

SOURCE

ORGANISM

Magnaporthe grisea.  
 Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;  
 Sordariomycetes incertae sedis; Magnaportheaceae; Magnaporthe.

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Construction and sequence analysis of an appressorium stage cDNA library in the rice blast fungus, Magnaporthe grisea  
 Unpublished (2001)  
 Contact: Ralph A. Dean  
 Fungal Genomics Laboratory  
 North Carolina State University  
 Campus Box 7251, Raleigh, NC 27695, USA  
 Tel: 919-513-0020  
 Fax: 919-513-0024  
 Email: ralph.dean@ncsu.edu

Seq primer: T3 primer (AATTAACCTCACTAAGG).

Location/Qualifiers

1..635  
 /organism="Magnaporthe grisea"  
 /strain="70-15"  
 /db\_xref="taxon:148305"  
 /clone="mgct006xe06f"  
 /clone\_1lb="Magnaporthe grisea Appressorium Stage cDNA"  
 /dev\_stage="germinated conidia on appressorium-inductive surface"  
 /note="Vector: pBluescript SK(+) Vector; site\_1: EcoRI; site\_2: XhoI. The appressorium formation-specific cDNA library was constructed from conidia germinated for 5-8 hr on an inductive surface. The library contains over 55,000 clones with average insert size of 1.5 kbp."

BASE COUNT 137 a 187 c 189 g 122 t

ORIGIN

## Query Match

Best Local Similarity 13.8%; Score 251.6; DB 12; Length 635;  
 Pred. No. 8.3e-42;

Matches 375; Conservative 0; Mismatches 184; Indels 3; Gaps 1;

449 GCGAAGATCAAAATATCTGCGGCTGCGCAATTCGCAATTCGCTGCGACATGAC 508  
 74 GCGAAGATCAAAATATCTGCGGCTGCGCAATTCGCAATTCGCTGCGACATGAC 133  
 509 GCGAGCTGTCCGATGACAGCTGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 565  
 134 GCGAGCTGTCCGATGACAGCTGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 193  
 566 GCGAGCTGTCCGATGACAGCTGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 625  
 194 GCGAGCTGTCCGATGACAGCTGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 685  
 626 GCTACATGCGAGTTTGTCTTCACACAGCGGTGAGCGCAAGTGTGAGAGCTCACTG 745  
 254 ATCTGCGGCAAGACCTGAGCAACACAGCGCGGTGAGCGCAAGTGTGAGAGCTTT 805  
 686 GCGTCTCAACAAGATGCTGACAGCGCTGTGCTGAGAGCGCGCTTCTGATGATGAC 865  
 314 GAGCGCTATGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 925  
 746 ATGACACACTTGGCGCTCAACAGCGCGATCAATCGCGCGAGGAGCGGTGCGAGAG 985  
 374 ATCCACAATTTGCGAGGTGGAGCGGCAAGATCATTTGGGCGAGGCGCGCCACAG 1045

GenCore version 5.1.4-p5\_4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 15, 2003, 15:59:49 ; Search time 2406.97 Seconds  
(without alignments)  
12286.366 Million cell updates/sec

Title: US-10-028-245-1  
Perfect score: 1826  
Sequence: 1 gtgcaccacgcgtcgtc.....atcaaaaaaaaaaaaaaa 1826

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues  
Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

EST:\*  
1: em\_estba:\*  
2: em\_estchun:\*  
3: em\_estin:\*  
4: em\_estmu:\*  
5: em\_estrov:\*  
6: em\_estrpl:\*  
7: em\_estro:\*  
8: em\_hic:\*  
9: gb\_est1:\*  
10: gb\_est2:\*  
11: gb\_hic:\*  
12: gb\_est3:\*  
13: gb\_est4:\*  
14: gb\_est5:\*  
15: em\_estfun:\*  
16: em\_estom:\*  
17: gb\_gss:\*  
18: em\_gss\_hum:\*  
19: em\_gss\_inv:\*  
20: em\_gss\_pln:\*  
21: em\_gss\_vrt:\*  
22: em\_gss\_fun:\*  
23: em\_gss\_mam:\*  
24: em\_gss\_mus:\*  
25: em\_gss\_other:\*  
26: em\_gss\_pro:\*  
27: em\_gss\_rod:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	521.4	28.6	658	13	BM077293 TREST-A09
2	251.6	13.8	635	12	BM810475 mgct006xe
3	109.8	6.0	317	14	BM865676 mgct003xc
4	78.8	4.3	997	17	CNS0057E
5	72.4	4.0	425	10	AW282705 LGL_292_H
6	70.4	3.9	1101	17	CNS017RP

7	67.2	3.7	987	17	CNS00418	AL06537 Drosophila
8	64.2	3.5	1101	17	CNS0181N	AL108773 Drosophila
9	56	3.1	1260	17	AO788019	AO788019 HS_3100_A
10	55.8	3.1	652	12	BM858164	BM858164 1024056B0
11	55.8	3.1	693	12	BM850114	BM850114 1024028A1
12	55.8	3.1	693	12	BM850115	BM850115 1024028A1
13	55.8	3.1	720	17	AG085797	AG085797 Pan trogl
14	55.8	3.1	1201	17	CNS0166G	AL106354 Drosophila
15	55	3.0	553	13	BM639114	BM639114 170006875
16	53.4	2.9	808	17	AG044302	AG044302 Pan trogl
17	52.6	2.9	885	17	CNS018H7	AL109333 Drosophila
18	52.4	2.9	530	12	BM904917	BM904917 Tatr1135G
19	52.2	2.9	1143	17	CNS012EH	AL174194 Tetradon
20	52.2	2.9	1996	9	AL697315	AL697315 AL697315
21	52	2.8	472	17	CNS0130K	AL103118 Drosophila
22	52	2.8	669	10	AV727802	AV727802 AV727802
23	52	2.8	959	17	CNS008M2	AL052079 Drosophila
24	51.8	2.8	982	13	BM606598	BM606598 170006870
25	51.6	2.8	681	17	CNS02EOD	AL193590 Tetradon
26	51.4	2.8	671	12	BM850116	BM850116 1024028A1
27	51.4	2.8	1067	17	CNS00C08	AL058564 Drosophila
28	51.4	2.8	1101	17	CNS006BP	AL064052 Drosophila
29	51	2.8	398	17	BM333647	BM333647 CH230-152
30	51	2.8	1101	17	CNS00F86	AL070460 Drosophila
31	50.8	2.8	416	12	BF435469	BF435469 nact1c12.
32	50.8	2.8	1101	17	CNS00ADP	AL056507 Drosophila
33	50.6	2.8	477	10	BE552639	BE552639 946083A01
34	50.4	2.8	656	13	BM416782	BM416782 953003D12
35	50.4	2.8	984	17	AG069149	AG069149 Pan trogl
36	50.2	2.7	604	10	AW438120	AW438120 707074A08
37	50.2	2.7	1183	17	CNS016CR	AL106581 Drosophila
38	50	2.7	381	10	BE552737	BE552737 946085A09
39	50	2.7	392	10	AW244923	AW244923 687016E08
40	50	2.7	524	9	AT664917	AT664917 605004C10
41	50	2.7	576	9	AT820133	AT820133 605086D05
42	50	2.7	582	9	AT173769	AT173769 606041A06
43	50	2.7	595	10	BE129978	BE129978 945034C01
44	50	2.7	600	10	BE238658	BE238658 946006E04
45	50	2.7	919	17	CNS006S5	AL065856 Drosophila

# ALIGNMENTS

RESULT 1	BM077293	658 bp	mRNA	linear	EST_05-FEB-2002
LOCUS	TREST-A0970	Hypocrea jecorina	CDNA	clone Tr-A0970 5'	
DEFINITION	BM077293	similar to cellulase, mRNA sequence.			
ACCESSION	BM077293.1	GI:18498475			
VERSION					
KEYWORDS					
SOURCE					
ORGANISM					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
MEDLINE					
COMMENT					

Chambergo, F.S., Bonaccorsi, E.D., Ferreira, A.J.S., Ramos, A.S.P.,  
Ferreira, J.R.Jr., Abrahao-Neto, J., Farah, J.P.S. and El-Dorri, H.  
Elucidation of the metabolic fate of glucose in the filamentous  
fungus *Trichoderma reesei* using expressed sequence tag (EST)  
analysis and cDNA microarrays  
J. Biol. Chem. 277 (16), 13983-13988 (2002)  
21950703  
Contact: El-Dorri, Hamza  
Department of Biochemistry  
Institute of Chemistry, University of Sao Paulo  
Avenida Professor Lineu Prestes, 748, Sao Paulo, SP, 05508-900,  
BRASIL  
Tel: (55) 11-3813848  
Fax: (55) 11-3813848  
Email: dorri@iq.usp.br  
PCR Primers

```
RESULT 15
US-09-910-186A-9
; Sequence 9, Application US/09910186A
; Publication No. US2003009025A1
; GENERAL INFORMATION:
; APPLICANT: U.S. Army Medical Research & Material Command
; TITLE OF INVENTION: RECOMBINANT VACCINE AGAINST BOTULINUM
; TITLE OF INVENTION: NEUROTOXIN
; FILE REFERENCE: A3626-A 067252.0107
; CURRENT APPLICATION NUMBER: US/09/910, 186A
; CURRENT FILING DATE: 2001-07-20
; PRIOR APPLICATION NUMBER: PCT/US00/12890
; PRIOR FILING DATE: 2000-05-12
; PRIOR APPLICATION NUMBER: 09/611,419
; PRIOR FILING DATE: 2000-07-06
; PRIOR APPLICATION NUMBER: 60/133,865
; PRIOR FILING DATE: 1999-05-12
; PRIOR APPLICATION NUMBER: 60/133,866
; PRIOR FILING DATE: 1999-05-12
; PRIOR APPLICATION NUMBER: 60/133,867
; PRIOR FILING DATE: 1999-05-12
; PRIOR APPLICATION NUMBER: 60/133,868
; PRIOR FILING DATE: 1999-05-12
; PRIOR APPLICATION NUMBER: 60/133,869
; PRIOR FILING DATE: 1999-05-12
; PRIOR APPLICATION NUMBER: 60/133,873
; PRIOR FILING DATE: 1999-05-12
; PRIOR APPLICATION NUMBER: 08/123,975
; PRIOR FILING DATE: 1993-09-21
; NUMBER OF SEQ ID NOS: 34
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 9
; LENGTH: 1371
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Synthetic construct based on Clostridium botulinum
; NAME/KEY: CDS
; LOCATION: (10)...(1359)
US-09-910-186A-9
```

```
Query Match          2.5%; Score 45.2; DB 9; Length 1371;
Best Local Similarity 49.6%; Pred. No. 0.0033;
Matches 116; Conservative 0; Mismatches 118; Indels 0; Gaps 0;

QY 642 TCCTCAACACGCGTGAGGAGGAGCTGACGAGCTCAACTGGGCTCTACACAAGG 701
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 29 TCCTCTCTACACCACTCCCTGTTGAAGACATCATCAAGAGTACTTCAACAACA 88
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 702 TCGTCACGCGCTGCTGAGACAGGCGCCCTACTGATGATGATGACATGACACTTTGCC 761
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 89 TCAACGACTCCAGATCTCTGCTGCAAGAACCTTAAGAACACCTGTGCGACACTCCG 148
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 762 GCTACACGCGCGCATCATGCGGAGGAGGCGTGTGAGACATCTTGTGCACTCT 821
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 149 GTTACACGCGCGAGGTCTCCGAGAGGAGTACGTCACGTGACCAATCTTCCCATTCG 208
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 822 GGGTCAGATCGCAAGTACTACGAGACAAGCAAGATCATCTTTGGCCTGA 875
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 209 ACTTCAAGCTGGGTCTCTCGGTGAGAGACAGAGGTAAAGTCAATGTCATCCACCA 262
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
```

Search completed: May 15, 2003, 16:06:16  
Job time : 249.351 secs

NUMBER OF SEQ ID NOS: 8  
 SOFTWARE: Patentin Ver. 2.1  
 SEQ ID NO 5  
 LENGTH: 905  
 TYPE: DNA  
 ORGANISM: Streptomyces albidoflavus  
 US-09-748-033-5

Query Match 2.5% Score 45.6; DB 10; Length 905;  
 Best Local Similarity 45.7%; Pred. No. 0.002; Indels 0; Gaps 0;  
 Matches 159; Conservative 0; Mismatches 189;

QY 518 CCGAGTACACGCTCTGTGCCCCCTGCTAGCTACAAAGAGAGATGGCGCCGAC 577  
 DB 13 CCGAGTACACGCTCTGTGCCCCCTGCTAGCTACAAAGAGAGATGGCGCCGAC 72  
 QY 578 ATGAGCATTTTGGCCCAAGACGAGCGCTTTCATGCTTTTGGCATTCGCTACATGGCAG 637  
 DB 73 CTGCGCGAGCTGCGCGAGCGCTTACGAGATCATGCGCGCTTTCGCGCGAGCGCC 132  
 QY 638 TTGTCTCAACACAGGCTGAGCGGAGCTGAGAGCTCAAGTGGGCTCTCAAC 697  
 DB 133 AACGCGGCGAGATCACTTACCTTCGAGCTCGGTGCGGCTCGCGCTACACCGAG 192  
 QY 698 AAGTCTCAACGCTGTCTGAGAGCGGCGCTTACATGATGATGACCAACTTT 757  
 DB 193 CAGTTCGCGCGAGCTGCGCGAGCGGAGCGGAGCGGAGTGTGATCATCTGCTG 252  
 QY 758 GCGCGTCAACGCGGCGCTTACATGCGGAGGAGCGGCTGTGCGAGCATCTTTGCGAC 817  
 DB 253 GCGCGGAGAAAGGCGCGGCTGCGGCTTACGAGAGCGGCTTCCCGCGCGGCGAC 312  
 QY 818 CTCTGCTCAGATCGCAAGTACTAGAGAGACAGCAAGATCATC 865  
 DB 313 AGCACTACGCGCTGATGAGAGATAGCGCTTGCAGCGGCTGCACATC 360

# RESULT 14

US-09-988-462-2  
 Sequence 2: Application US/09988462  
 Publication No. US20030046726A1

GENERAL INFORMATION:  
 APPLICANT: Koziel, Michael G.  
 Desai, Nalin M.  
 Lewis, Kelly S.  
 Kramer, Vance C.  
 Warren, Gregory W.  
 Evola, Stephen V.  
 Crossland, Lytle D.  
 Wright, Martha S.  
 Merlin, Ellis J.  
 Lunnis, Karen L.

TITLE OF INVENTION: SYNTHETIC DNA SEQUENCE HAVING ENHANCED  
 INSECTICIDAL ACTIVITY IN MAIZE

NUMBER OF SEQUENCES: 94  
 CORRESPONDENCE ADDRESS:  
 ADDRESS: Syngenta Biotechnology, Inc.  
 STREET: 3054 Cornwells Road  
 CITY: Research Triangle Park  
 STATE: NC  
 COUNTRY: USA  
 ZIP: 27709

COMPUTER READABLE FORM:  
 MEDIUM TYPE: floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/09/988,462  
 FILING DATE: 20-NO. US20030046726A1-2001

CLASSIFICATION: <Unknown>  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 09/547,422

FILING DATE: 11-APR-2000  
 APPLICATION NUMBER: US 08/459,504  
 FILING DATE: 02-JUN-1995  
 APPLICATION NUMBER: US 07/951,715  
 FILING DATE: 25-SEP-1992  
 APPLICATION NUMBER: US 07/772,027  
 FILING DATE: 04-OCT-1991  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Meigs, J. Timothy  
 REGISTRATION NUMBER: 38,241  
 REFERENCE/DOCKET NUMBER: S-18051

TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (919)541-8587  
 TELEFAX: (919)541-8689  
 INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:  
 LENGTH: 3468 base pairs  
 TYPE: nucleic acid  
 STRANDEDNESS: single  
 TOPOLOGY: linear

MOLECULE TYPE: other nucleic acid  
 DESCRIPTION: /desc = "Synthetic DNA"

HYPOTHETICAL: NO  
 FEATURE:  
 NAME/KEY: misc-feature  
 LOCATION: 1..3468  
 OTHER INFORMATION: /product= "Full-length pure maize  
 optimized synthetic Bt"

/note="disclosed in Figure 3 as syn17.mze"

SEQUENCE DESCRIPTION: SEQ ID NO: 2:  
 US-09-988-462-2

Query Match 2.5% Score 45.4; DB 9; Length 3468;  
 Best Local Similarity 44.5%; Pred. No. 0.0049;  
 Matches 230; Conservative 0; Mismatches 281; Indels 6; Gaps 1;

QY 642 TCCTCAACACAGCGGTGAGCGGAGCTGAGAGCTCAACTGGGCTCTTACAAAG 701  
 DB 2657 TCCTGAACAGCGGTGAGCGGAGCTGAGAGCTGAGAGCTCAACTGGGCTCTTACAAAG 716  
 QY 702 TCCTCAACAGCGGTGAGCGGAGCTGAGAGCTGAGAGCTCAACTGGGCTCTTACAAAG 761  
 DB 2717 CCAGCAAGCGGTGAGCGGAGCTGAGAGCTGAGAGCTCAACTGGGCTCTTACAAAG 776  
 QY 762 GCTCAACAGCGGTGAGCGGAGCTGAGAGCTGAGAGCTCAACTGGGCTCTTACAAAG 821  
 DB 2777 GCTGAACAGCGGTGAGCGGAGCTGAGAGCTGAGAGCTCAACTGGGCTCTTACAAAG 836  
 QY 822 GGTTCAGATCGCAAGTACTAGAGAGCAACAGCAAGATCATCTTGGCCTGATGAAG 881  
 DB 2837 AGCAGCGCGCGAGATGATCAAGAGCGGAGCTGAGAGCTCAACTGGGCTCTTACAAAG 896  
 QY 882 AGCGGAGCGGTGAGCGGAGCTGAGAGCTGAGAGCTCAACTGGGCTCTTACAAAG 941  
 DB 2897 TGAAGGCGCGGTGAGCGGAGCTGAGAGCTGAGAGCTCAACTGGGCTCTTACAAAG 956  
 QY 942 TCGCAAGCGGTGAGCGGAGCTGAGAGCTGAGAGCTCAACTGGGCTCTTACAAAG 1001  
 DB 2957 AGTGGAGCGGTGAGCGGAGCTGAGAGCTGAGAGCTCAACTGGGCTCTTACAAAG 1016  
 QY 1002 TCGAGAGTATGTCATCAGCGGAGCGGAGCTGAGAGCTCAACTGGGCTCTTACAAAG 1061  
 DB 3017 GGTGAACCGGTGAGCGGAGCTGAGAGCTGAGAGCTCAACTGGGCTCTTACAAAG 1076  
 QY 1062 GAAGCAGCGGTGAGCGGAGCTGAGAGCTGAGAGCTCAACTGGGCTCTTACAAAG 1115  
 DB 3077 ACAACAGCGGTGAGCGGAGCTGAGAGCTGAGAGCTCAACTGGGCTCTTACAAAG 1136  
 QY 1116 CCGGTCGAGCGGAGCTGAGAGCTGAGAGCTCAACTGGGCTCTTACAAAG 1152  
 DB 3137 CCGTGAAGCGGTGAGCGGAGCTGAGAGCTGAGAGCTCAACTGGGCTCTTACAAAG 1173







```

1 RESULT 4
2 US-09-927-827-26
3 : Sequence 26. Application US/09927827
4 : Publication NO. US20030036176A1
5 :
6 : GENERAL INFORMATION:
7 :
8 : APPLICANT: Bower, Stanley G.
9 : APPLICANT: Ramseler, Thomas M.
10 :
11 : TITLE OF INVENTION: Directed Genetic Engineering of Xanthomonas campestris
12 :
13 : FILE REFERENCE: 38-10(15824)B
14 :
15 : CURRENT APPLICATION NUMBER: US/09/927,827
16 :
17 : CURRENT FILING DATE: 2001-08-10
18 :
19 : PRIOR APPLICATION NUMBER: US 60/279,493
20 :
21 : PRIOR FILING DATE: 2001-03-28
22 :
23 : NUMBER OF SEQ ID NOS: 69
24 :
25 : SEQ ID NO 26
26 :
27 : LENGTH: 2040
28 :
29 : TYPE: DNA
30 :
31 : ORGANISM: Xanthomonas campestris
32 :
33 : FEATURE:

```

[illegible]

```

Query Match 3.6% Score 65.8; DB 9; Length 2040;
Best Local Similarity 53.3%; Pred. No. 1,8e-09;
Matches 139; Conservative 0; Mismatches 122; Indels 0; Gaps 0;

OY 723 CGGGGCCCTACTGTCATGATTGCATATGACACAACTTTGCCCGCTACAAAGCGGGCATCATCG 782
Db 1293 CGGGCATGATATCTGGTGGCTGGATATTCACAACTACAGCAAGTACTGATTAAATGG 1352
OY 783 GCCAGGAGAGCGCTGTCCGACGACATCTTGTCCACTCTGGGGTCCAGATCGCAAGTACT 842
Db 1353 GTGGGCGCGAAGTGGCGGCTCGCCACTTTGGCCGATCTGGGGGGCGGCGCTGGCGTGATCT 1412
OY 843 ACGAGGACACAGCAGCAAGATCATCTTTGGCGCTGTACAGAGCGGACGACACTCGACATTG 902
Db 1413 TCACAGGAGTAAAGCCCGTGATCTTGGGGCTATGAATGAGGCCAAACATCTTGCCA 1472
OY 903 AGATCTGGGGCCGACAGCGTGCACAAAGGTCGTACTGCGATCCGAAAGCGGGCGCCACT 962
Db 1473 GCGAATGGGGCCGCTGCACGCGCAGGCGAGCATGATGCATCGCTGCCACCGGCGCCACA 1532
OY 963 CGCAGATGATGCTCCTCGCCG 983
Db 1533 ATCTGATCTCTGTTACCGGCG 1553

RESULT 5
US-09-790-399-7
Sequence 79, Application US/09790399
Patent No. US20020038000A1
GENERAL INFORMATION:
APPLICANT: Gold, Larry
APPLICANT: Tuerk, Craig
APPLICANT: Pribnow, David
APPLICANT: Smith, Jonathan D.
TITLE OF INVENTION: Systematic Polypeptide Evolution by Reverse Translation
FILE REFERENCE: NEK02/CI-CON2
CURRENT FILING DATE: US/09/790,399
CURRENT FILING DATE: 2001-02-22
PRIOR APPLICATION NUMBER: 09/197,649
PRIOR FILING DATE: 1998-11-23
PRIOR APPLICATION NUMBER: 07/829,461
PRIOR FILING DATE: 1992-01-31
PRIOR APPLICATION NUMBER: 07/739,055
PRIOR FILING DATE: 1991-08-01
PRIOR APPLICATION NUMBER: 07/561,968
PRIOR FILING DATE: 1990-08-02
NUMBER OF SEQ ID NOS: 26
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 7
LENGTH: 390
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: Sequence
OTHER INFORMATION: having a 120 repeat of ACG flanked by fixed
OTHER INFORMATION: fragments having NcoI restriction sites.
US-09-790-399-7

```

```

Db 358 CCTGCGATGGCAGTACCTGCTCAACAACTTTGGCGGCAATTTGATTCACGAGC 417
OY 683 TGGGGCTCTACAACAAGCTGTCACAGCCCTGCTCGAGAGCGGGCCCTGATGAT 742
Db 418 ATTTCAGATATGATGACGCTTTGTCAGGGGCTGCTCTCGGGCCGATGATGATG 477
OY 743 GACATGCAACTTTGCCCCGTACACAGCGCGGATCATCGGCGAGGAGGCGTTCGAC 802
Db 478 GACATTCACAAATTTGCTCGATGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 537
OY 803 GACATCTTTGCGACCTGCGGCTGCGGAGTATGCAAGTATGCAAGTATGCAAGTATG 862
Db 538 GCTCAATTCAGAGCCTTTGCTGCGAGTATGCAAGTATGCAAGTATGCAAGTATG 597
OY 863 ATCTTTGCGCTGATGAAGAGAGCGGCGGAGCTGCAATTTGATGATGATGATGATG 922
Db 598 TGGTGGGCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 657
OY 923 CAAAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 982
Db 658 CAAAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 717
OY 983 GGAACCAACTTTGCGACGCTGAGAGAGTATGCTGCTGCTGCTGCTGCTGCTGCTG 1042
Db 718 GGAATATATTTGCGAATTTGCTGCGGCTTTTCAATTCAGATGAGAGAGAGAGAG 777
OY 1043 AGATTTACGAACCGGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1102
Db 778 CAGATCAGAACCGGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 837
OY 1103 GACATCAACAATTCGCGGCTGCGAGCGGAGTGCACACAGACAAAC---GTCGAGCCTTC 1159
Db 838 GACATCAACAATTCGCGGCTGCGAGCGGAGTGCACACAGACAAAC---GTCGAGCCTTC 897
OY 1160 AAGCATTTGCGGAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1219
Db 898 TCTCCGCTTGGCAATTTGCTGCGGCTTTTCAATTCAGATGAGAGAGAGAGAGAG 957
OY 1220 GCGTCATGAGAGCTTGTGATGATGATGATGATGATGATGATGATGATGATGATG 1279
Db 958 GGTGCGCAAGCTGATGCTGCTGATGATGATGATGATGATGATGATGATGATGATG 1017
OY 1280 AACAGCAGCTTCAATGCTTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1339
Db 1018 AACTCAATATCTATCTGCTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1077
OY 1340 ATCTTGACTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1385
Db 1078 GTCTGACGGAACACGAGTACGATGATGATGATGATGATGATGATGATGATGATG 1123

RESULT 2
US-09-927-827-20/c
; Sequence 20, Application US/09927827
; Publication No. US20030036176A1
; GENERAL INFORMATION:
; APPLICANT: Bower, Stanley G.
; APPLICANT: Ramseler, Thomas M.
; TITLE OF INVENTION: Directed Genetic Engineering of Xanthomonas campestris
; FILE REFERENCE: 38-10(15824)B
; CURRENT FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: US 60/279,493
; NUMBER OF SEQ ID NOS: 69
; SEQ ID NO 20
; LENGTH: 2957
; TYPE: DNA
; ORGANISM: Xanthomonas campestris
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1001)..(1957)

```

```

US-09-927-827-20
Query Match 7.2%; Score 132; DB 9; Length 2957;
Best Local Similarity 51.5%; Pred. No. 8.2e-30;
Matches 384; Conservative 0; Mismatches 350; Indels 12; Gaps 3;

OY 643 CCTCAACAACAGGTTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 702
Db 1793 CCTGAGCAACCTTCATGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1734
OY 703 CGTCAAGCCTTGTCTCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 762
Db 1733 CGTCAAGCCTTGTCTCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1674
OY 763 CTACAGCGGCGGATCATGCTGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 822
Db 1673 GTACAGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1614
OY 823 GGTCCAGATCGCAAGTATGATGATGATGATGATGATGATGATGATGATGATGATG 882
Db 1613 GCGCGCGCTGCGCTGAGATGATGATGATGATGATGATGATGATGATGATGATGAT 1554
OY 883 GCGCGAGAGCTGCAATGATGATGATGATGATGATGATGATGATGATGATGATGAT 942
Db 1553 GCGCGAGAGCTGCAATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1494
OY 943 CCGAAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1002
Db 1493 CCGAAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1437
OY 1003 CGAGACGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1059
Db 1436 GCACAGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1377
OY 1060 TGAAGACGAGATTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1119
Db 1376 GAGCCCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1317
OY 1120 GTGCGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 1173
Db 1316 CACCAAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1257
OY 1174 CTGCGTGAAGCAGAACAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1233
Db 1256 CTGCGTGAAGCAGAACAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1197
OY 1234 TTGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1293
Db 1196 GGTCTGCAACGAGCGGCTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1137
OY 1294 CATTGGCTTTGTGGGCTGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1353
Db 1136 GCTGGGCTGAGACCTGATGATGATGATGATGATGATGATGATGATGATGATGATG 1077
OY 1354 TCCCTCGGCAAGCGGCGGCAACTACA 1379
Db 1076 GCAAGCGGCGGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1051

RESULT 3
US-09-927-827-25
; Sequence 25, Application US/09927827
; Publication No. US20030036176A1
; GENERAL INFORMATION:
; APPLICANT: Bower, Stanley G.
; APPLICANT: Ramseler, Thomas M.
; TITLE OF INVENTION: Directed Genetic Engineering of Xanthomonas campestris
; FILE REFERENCE: 38-10(15824)B
; CURRENT FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: US 60/279,493
; NUMBER OF SEQ ID NOS: 69

```

GenCore version 5.1.4-p5-4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 15, 2003, 13:49:29 ; Search time 221.351 Seconds

(without alignments)  
10641.144 Million cell updates/sec

Title: US-10-028-245-1

Perfect score: 1826  
Sequence: 1 gtcgaccacgcgtccgtc.....atcaaaaaaaaaaaaaa 1826

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 810007 seqs, 644969091 residues

Total number of hits satisfying chosen parameters: 1620014

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published\_Applications\_NA.\*  
1: /cgn2\_6/ptodata/2/pubpna/US07\_PUBCOMB.seq.\*  
2: /cgn2\_6/ptodata/2/pubpna/PCT\_NEW\_PUB.seq.\*  
3: /cgn2\_6/ptodata/2/pubpna/US06\_NEW\_PUB.seq.\*  
4: /cgn2\_6/ptodata/2/pubpna/US07\_NEW\_PUB.seq.\*  
5: /cgn2\_6/ptodata/2/pubpna/US07\_PUBCOMB.seq.\*  
6: /cgn2\_6/ptodata/2/pubpna/US07\_PUBCOMB.seq.\*  
7: /cgn2\_6/ptodata/2/pubpna/US08\_NEW\_PUB.seq.\*  
8: /cgn2\_6/ptodata/2/pubpna/US08\_PUBCOMB.seq.\*  
9: /cgn2\_6/ptodata/2/pubpna/US09\_NEW\_PUB.seq.\*  
10: /cgn2\_6/ptodata/2/pubpna/US09\_PUBCOMB.seq.\*  
11: /cgn2\_6/ptodata/2/pubpna/US10\_NEW\_PUB.seq.\*  
12: /cgn2\_6/ptodata/2/pubpna/US10\_PUBCOMB.seq.\*  
13: /cgn2\_6/ptodata/2/pubpna/US60\_NEW\_PUB.seq.\*  
14: /cgn2\_6/ptodata/2/pubpna/US60\_PUBCOMB.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	275.6	15.1	1155	US-09-916-494A-15	Sequence 15, Appl
2	132	7.2	2957	US-09-927-827-20	Sequence 20, Appl
3	75.8	4.2	3002	US-09-927-827-25	Sequence 25, Appl
4	65.8	3.6	2040	US-09-927-827-26	Sequence 26, Appl
5	55.6	3.0	390	US-09-790-399-7	Sequence 7, Appl
6	51.8	2.8	1185	US-09-887-576-784	Sequence 784, App
7	47	2.6	1294	US-09-748-033-2	Sequence 2, Appl
8	46.8	2.6	520	US-10-184-644-332	Sequence 332, App
9	46.8	2.6	520	US-10-184-634-332	Sequence 332, App
10	46.4	2.5	2903	US-09-984-827-138	Sequence 138, App
11	46	2.5	1347	US-09-350-756-6	Sequence 6, Appl
12	45.8	2.5	905	US-09-748-033-7	Sequence 7, Appl
13	45.6	2.5	905	US-09-748-033-5	Sequence 5, Appl
14	45.4	2.5	3468	US-09-988-462-2	Sequence 2, Appl
15	45.2	2.5	1371	US-09-910-186A-9	Sequence 9, Appl
16	45.2	2.5	1371	US-09-350-756-3	Sequence 3, Appl
17	45.2	2.5	1734	US-09-738-626-1062	Sequence 1062, Ap
18	44.4	2.4	1098	US-09-815-242-7874	Sequence 7874, Ap
19	44.4	2.4	1368	US-09-910-186A-17	Sequence 17, Appl

20	44.4	2.4	1593	10	US-09-997-664-102	Sequence 102, App
21	44.4	2.4	6491	10	US-09-997-664-1	Sequence 1, Appl
22	43.8	2.4	492	9	US-09-925-299-586	Sequence 586, App
23	43.8	2.4	492	10	US-09-925-299-586	Sequence 586, App
24	43.8	2.4	1107	10	US-09-748-033-6	Sequence 6, Appl
25	43.8	2.4	2712	10	US-09-748-033-4	Sequence 4, Appl
26	43.8	2.4	3468	9	US-09-988-462-4	Sequence 4, Appl
27	43.8	2.4	53522	9	US-09-904-968A-1	Sequence 1, Appl
28	43.4	2.4	726	10	US-09-864-761-19707	Sequence 19707, A
29	43.4	2.4	1030	10	US-09-804-682-16	Sequence 16, Appl
30	43.4	2.4	1951	10	US-09-864-761-2926	Sequence 2926, Ap
31	43.2	2.4	475	10	US-09-864-761-1361	Sequence 1361, Ap
32	43.2	2.4	512	10	US-09-864-761-18121	Sequence 18121, A
33	43.2	2.4	32152	9	US-09-764-872-518	Sequence 518, App
34	43.2	2.4	32152	9	US-10-072-349-328	Sequence 328, App
35	43.2	2.4	32152	10	US-09-764-855-338	Sequence 338, App
36	43.2	2.4	1503841	9	US-09-946-807-1	Sequence 1, Appl
37	43.2	2.4	1503841	10	US-09-795-668-1	Sequence 1, Appl
38	43.2	2.4	1503841	10	US-09-795-668-1	Sequence 1, Appl
39	43	2.4	6442	9	US-09-950-335A-11	Sequence 11, Appl
40	42.8	2.3	267	10	US-09-864-761-27984	Sequence 27984, A
41	42.8	2.3	466	9	US-09-918-995-776	Sequence 776, App
42	42.8	2.3	1184	9	US-10-123-155-394	Sequence 394, App
43	42.8	2.3	3809	12	US-10-001-870-68	Sequence 68, Appl
44	42.4	2.3	485	9	US-09-918-995-1001	Sequence 1001, Ap
45	42.4	2.3	19000	9	US-10-109-084-3	Sequence 3, Appl

#### ALIGNMENTS

RESULT 1  
US-09-916-494A-15  
Sequence 15, Application US/09916494A  
Patent No. US20020164774A1  
GENERAL INFORMATION:  
APPLICANT: Fowler, Timothy  
APPLICANT: Clarkson, Kathleen A.  
APPLICANT: Ward, Michael  
APPLICANT: Collier, Katherine D.  
TITLE OF INVENTION: Method and Compositions for Treating  
TITLE OF INVENTION: Cellulose Containing Fabrics Using Truncated Cellulase  
FILE REFERENCE: GC226-C4  
CURRENT APPLICATION NUMBER: US/09/916,494A  
CURRENT FILING DATE: 2000-06-14  
PRIOR APPLICATION NUMBER: US 08/382,452  
PRIOR FILING DATE: 1995-02-01  
PRIOR APPLICATION NUMBER: US 08/169,948  
PRIOR FILING DATE: 1993-12-17  
NUMBER OF SEQ ID NOS: 43  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 15  
LENGTH: 1155  
TYPE: DNA  
ORGANISM: Trichoderma longibrachiatum  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (1)...(56)  
NAME/KEY: CDS  
LOCATION: (231)...(1155)  
US-09-916-494A-15  
Query Match 15.1%; Score 275.6; DB 9; Length 1155;  
Best Local Similarity 59.2%; Pred. No. 2, 2e-74;  
Matches 489; Conservative 0; Mismatches 334; Indels 3; Gaps 1;  
QY 563 GATGGGCGCGCGAGATTTTGGCGAGAGAGCGCCGCAACGCTTCGATA 622  
DB 298 GATGGGCGCGCGAGATTTTGGCGAGAGAGCGCCGCAACGCTTCGATA 357  
QY 623 TCCGCTACATGCGAGCTTCTCTCAACACAGCGTGAGCGCAAGCTGAGAGCTCAAC 682

THIS PAGE BLANK (USPTO)

THIS PAGE BLANK (USPTO)

```

Db 777 GAGATCAACAACCTACCTGGCCAAACAGGGGGGCGGCGGCAAGCAAGCTGAGCCAG 836
QY 758 GCCCGCTACAGGGGGGATCATCGGCCAGGGAGGGGTGTGCGACGACATCTTTGCGAC 817
Db 837 ATCAAGAACATCAAGCAGCGCTGGGGCAAGAACCCATCCCGAAGAACATCACCGGTAC 896
QY 818 CTCTGGGTCCAGATCCGAAAGTACTACGAGGACAGACAGACATCATCTTTGGCTGATG 877
Db 897 CGGTGGTGGCGCATGGCCGAGTTCCGGCTACCAATCAGGAGACCCCTGCCACGCTGAG 956
QY 878 AACGACCCGACGACCTCGACATTTGAGATCTGGCGGAGAGCGTCCAAAAGTCTGACT 937
Db 957 GACTTCGAGGAGAGTCTCTGAACACCATCAAGGAGAGACAGGCTACATGAGCAGCAGC 1016
QY 938 GCGATCCGAAGGCGGCGGCACT 962
Db 1017 CTGAGCAGGAGCGGCTGGCGGCT 1041

```

## RESULT 15

US-08-470-566B-39

Sequence 39, Application US/08470566B

Patent No. 5872212

GENERAL INFORMATION:

APPLICANT: Warren, Gregory W

APPLICANT: Koziel, Michael G

APPLICANT: Mullins, Martha A

APPLICANT: Nye, Gordon J

APPLICANT: Carr, Brian

APPLICANT: Desai, Nalin M

APPLICANT: Kostichka, N. Kristy

APPLICANT: Duck, Nicholas B

APPLICANT: Estruch, Juan J

TITLE OF INVENTION: No 5872212el Pesticidal Proteins and Strains

NUMBER OF SEQUENCES: 52

CORRESPONDENCE ADDRESS:

ADDRESSEE: No. 5872212artis Corporation

STREET: 3054 Cornwallis Road

CITY: Research Triangle Park

STATE: NC

COUNTRY: USA

ZIP: 27709

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent Release #1.0, Version #1.30B

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/470,566B

FILING DATE: 06-JUN-1995

CLASSIFICATION: 530

PRIORITY APPLICATION DATA:

APPLICATION NUMBER: US 08/463,483

FILING DATE: 05-JUN-1995

PRIORITY APPLICATION DATA:

APPLICATION NUMBER: US 08/314,594

FILING DATE: 09-SEP-1994

PRIORITY APPLICATION DATA:

APPLICATION NUMBER: US 08/218,018

FILING DATE: 23-MAR-1994

PRIORITY APPLICATION DATA:

APPLICATION NUMBER: US 08/037,057

FILING DATE: 25-MAR-1993

ATTORNEY/AGENT INFORMATION:

NAME: Meigs, J. Timothy

REGISTRATION NUMBER: 38,241

REFERENCE/DOCKET NUMBER: CGC1695/CIP3/DIV4 - SOLV4

TELEPHONE: 919-541-8587

TELEFAX: 919-541-8689

INFORMATION FOR SEQ ID NO: 39:

SEQUENCE CHARACTERISTICS:

```

? LENGTH: 1241 base pairs
? TYPE: nucleic acid
? STRANDEDNESS: single
? TOPOLOGY: linear
? MOLECULE TYPE: other nucleic acid
? DESCRIPTION: /desc = "Synthetic DNA"
? FEATURE:
? NAME/KEY: CDS
? LOCATION: 9..1238
? OTHER INFORMATION: /note= "Maize optimized DNA"
? OTHER INFORMATION: sequence encoding VIP2(a) with the Bacillus secretion sig
? OTHER INFORMATION: removed as contained in pIB5527"
US-08-470-566B-39

```

Query Match 2.9% Score 53.8; DB 2; Length 1241;  
 Best Local Similarity 46.2% Pred. No. 0.00075;  
 Matches 178; Conservative 0; Mismatches 207; Indels 0; Gaps 0;

```

QY 578 ATGAGCATTTCCGCCGAAAGAGAGAGGCGCTTCACGCTTTTCGATATCCGCTACATGGCAG 637.
Db 657 AAGAGCATATCAACGCGGAGGCGCCACAGCTGGGGCATGAAGAATACAGAGAGTGGGCC 716
QY 638 TTGTCTCTACACACACGCTGAGCGCAAGCTGAGCAGAGCTCAACTGGGCTCTACAC 697
Db 717 AAGGACCTGACCGACAGCCAGCGCGGAGGCGCTGAGCGCTACGCGCCGACGACTACAG 776
QY 698 AAGTGCTCAACGCGCTGTCTGAGAGGCGGCGCTACTGATGATTTGATGACACTTT 757
Db 777 GAGATCAACACTTACTGCGCCAAACAGGCGGCGCAAGAGAGAGTGGACGCCAG 836
QY 758 GCCCGCTACAGGGGGGATCATCGGCCAGGGAGGCGGTGTGCGACGACATCTTTGCGAC 817
Db 837 ATCAAGAACATCAAGCGCGGCGCTGGGCAAGAACCCATCCCGAAGAACATCACCGGTAC 896
QY 818 CTCTGGGTCCAGATCCGAAAGTACTACGAGGACAGACAGACATCATCTTTGGCTGATG 877
Db 897 CGGTGGTGGCGCATGGCCGAGTTCTGGCTACAGATCAGCAGCCCGCTGCCACGCTGAG 956
QY 878 AACGAGCGGACAGACCTCGACATTTGAGATCTGGGCGGAGAGCGCAAGAGTCTGACT 937
Db 957 GACTTCGAGGAGAGTCTCTGAACACCATCAAGGAGAGACAGGCTACATGAGCAGCAGC 1016
QY 938 GCGATCCGAAGGCGGCGGCACT 962
Db 1017 CTGAGCAGGAGCGGCTGGCGGCT 1041

```

Search completed: May 15, 2003, 13:49:06  
 Job time: 81.2978 secs

PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 08/218,018  
 FILING DATE: 23-MAR-1994  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 08/037,057  
 FILING DATE: 25-MAR-1993  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Meigs, J. Timothy  
 REGISTRATION NUMBER: 38,241  
 REFERENCE/DOCKET NUMBER: CGC1695/CIP3/DIV8 - SOLV4  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 919-541-8587  
 TELEFAX: 919-541-8689  
 INFORMATION FOR SEQ ID NO: 39:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 1241 base pairs  
 TYPE: nucleic acid  
 STRANDEDNESS: single  
 TOPOLOGY: linear  
 MOLECULE TYPE: other nucleic acid  
 DESCRIPTION: /desc = "Synthetic DNA"  
 FEATURE:  
 NAME/KEY: CDS  
 LOCATION: 9..1238  
 OTHER INFORMATION: /note= "Maize optimized DNA  
 OTHER INFORMATION: sequence encoding VIP2A(a) with the Bacillus secretion signal  
 OTHER INFORMATION: removed as contained in pcib5528"

Query Match 2.98; Score 53.8; DB 2; Length 1241;  
 Best Local Similarity 46.2%; Pred. No. 0.00075;  
 Matches 178; Conservative 0; Mismatches 207; Indels 0; Gaps 0;

578 ATGAAGCATTTGCGGCAAGAGGAGGCGCTCAAGCTTTTGGCATATCCGATACATGGCAG 637  
 657 AAGAGACATCAACGCGGAGGCGCCACAGCTGGGGCATGAAGACTACGAGAGTGGGCC 716  
 638 TTGTCTCAACAACAGCGGTGAGCGGAGCTGAGCAGCTCAACTGGGGCTCTTACAC 697  
 717 AAGAGCTGACCGACCGACCGGAGGCGCTGGAGCGCTAGCGCCGACGAGTACAG 776  
 698 AAGGTCTCAACGCGCTCTCGAGAGCGGCGCTACTGATGATGACATGCAACTTT 757  
 777 GAGATCAACACTGCTGCGCAACAGGCGGCGGCAACGAGAGCTGGAGCGCCAG 836  
 758 GCCGCTACACGCGGCGCATCATGCGGAGGAGCGGTGCGAGCATCTTTGTGAC 817  
 837 ATCAAGAACATCAAGCGGCGGCGGCGGCAAGAGCCCATCCCGAATCACCCTGTAC 896  
 818 CTCTGGGTCCAGATCGCAAGTACTACGAGACACAGCAAGATCATCTTTGGCTGATG 877  
 897 CGCTGGTGGCGCATGCGCGGAGTTCGCTACAGATCAGCGACCCCTGCCAGGCTAAG 956  
 878 AAGAGCGGAGACAGCTCGACATTTGAGTGGCGGAGAGCTGCCAAGAGTGTACT 937  
 957 GACTTCGAGAGGAGCTTCTGAACACCATTAAGAGAGGAGGAGGCTACATGAGACAGC 1016  
 938 GCGATCCGAAGGCGGCGGCGCCACT 962  
 1017 CTGAGCAGGAGCGGCGGCGGCGCT 1041

RESULT 14  
 US-08-471-046A-42  
 Sequence 42, Application US/08471046A  
 Patent No. 5866326

GENERAL INFORMATION:  
 APPLICANT: Warren, Gregory M  
 APPLICANT: Koziel, Michael G  
 APPLICANT: Mullins, Martha A  
 APPLICANT: Nye, Gordon J  
 APPLICANT: Carr, Brian

APPLICANT: Desai, Nalini M  
 APPLICANT: Kostichka, N. Kristy  
 APPLICANT: Duck, Nicholas B  
 APPLICANT: Estruch, Juan J  
 TITLE OF INVENTION: Method For Isolating Vegetative Insecticidal  
 TITLE OF INVENTION: Protein Genes  
 NUMBER OF SEQUENCES: 50  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: No. 5866326arlis Corporation  
 STREET: 3054 Cornwallis Road  
 CITY: Research Triangle Park  
 STATE: NC  
 COUNTRY: USA  
 ZIP: 27709  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: PatentIn Release #1.0, Version #1.30B  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/471,046A  
 FILING DATE: 06-JUN-1995  
 CLASSIFICATION: 435  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 08/463,483  
 FILING DATE: 05-JUN-1995  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 08/314,594  
 FILING DATE: 09-SEP-1994  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 08/218,018  
 FILING DATE: 23-MAR-1994  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 08/037,057  
 FILING DATE: 25-MAR-1993  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Meigs, J. Timothy  
 REGISTRATION NUMBER: 38,241  
 REFERENCE/DOCKET NUMBER: CGC1695/CIP3/DIV8 - SOLV4  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 919-541-8587  
 TELEFAX: 919-541-8689  
 INFORMATION FOR SEQ ID NO: 42:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 1241 base pairs  
 TYPE: nucleic acid  
 STRANDEDNESS: single  
 TOPOLOGY: linear  
 MOLECULE TYPE: other nucleic acid  
 DESCRIPTION: /desc = "Synthetic DNA"  
 FEATURE:  
 NAME/KEY: CDS  
 LOCATION: 9..1238  
 OTHER INFORMATION: /note= "Maize optimized DNA  
 OTHER INFORMATION: sequence encoding VIP2A(a) with the Bacillus secretion sig  
 OTHER INFORMATION: removed as contained in pcib5528"  
 OTHER INFORMATION: contained in pcib5528"

Query Match 2.98; Score 53.8; DB 2; Length 1241;  
 Best Local Similarity 46.2%; Pred. No. 0.00075;  
 Matches 178; Conservative 0; Mismatches 207; Indels 0; Gaps 0;

578 ATGAAGCATTTGCGGCAAGAGGAGGCGCTCAAGCTTTTGGCATATCCGATACATGGCAG 637  
 657 AAGAGACATCAACGCGGAGGCGCCACAGCTGGGGCATGAAGACTACGAGAGTGGGCC 716  
 638 TTGTCTCAACAACAGCGGTGAGCGGAGCTGAGCAGCTCAACTGGGGCTCTTACAC 697  
 717 AAGAGCTGACCGACCGACCGGAGGCGGCTGGAGCGCTAGCGCCGACGAGTACAG 776  
 698 AAGGTCTCAACGCGCTCTCGAGAGCGGCGGCGCTACTGATGATGACATGCAACTTT 757

Db 957 GACTTCGAGGACAGCTTCCTGAAACACATCAAGAGAGAGGAGGCTACATGAGACCCAGC 1016  
 Oy 938 GCGATCCGAAGAGCGCGGCCCACT 962  
 Db 1017 CTGAGCAGCGAGCGCTGCGCCCT 1041

## RESULT 12

US-08-463-483A-42

Sequence 42, Application US/08463483A

Patent No. 5849870

GENERAL INFORMATION:

APPLICANT: Warren, Gregory M

APPLICANT: Kozziel, Michael G

APPLICANT: Mullins, Martha A

APPLICANT: Nye, Gordon J

APPLICANT: Carr, Brian

APPLICANT: Desai, Nalin M

APPLICANT: Kostichka, N. Kristy

APPLICANT: Duck, Nicholas B

TITLE OF INVENTION: No. 5849870el Pesticidal Proteins and Strains

NUMBER OF SEQUENCES: 50

CORRESPONDENCE ADDRESS:

ADDRESSEE: CIBA-GEIGY Corporation

STREET: 7 Skyline Drive

CITY: Hawthorne

STATE: NY

COUNTRY: USA

ZIP: 10532

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent Release #1.0, Version #1.308

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/463,483A

FILING DATE:

CLASSIFICATION: 530

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/314,594

FILING DATE: 09-SEP-1994

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/218,018

FILING DATE: 23-MAR-1994

APPLICATION DATA:

APPLICATION NUMBER: US 08/037,057

FILING DATE: 25-MAR-1993

ATTORNEY/AGENT INFORMATION:

NAME: Spull, W. Murray

REGISTRATION NUMBER: 32,943

REFERENCE/DOCKET NUMBER: CGC 1695/CIP3

TELECOMMUNICATION INFORMATION:

TELEPHONE: 919-541-8615

INFORMATION FOR SEQ ID NO: 42:

SEQUENCE CHARACTERISTICS:

LENGTH: 1241 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

MOLECULE TYPE: other nucleic acid

DESCRIPTION: /desc = "Synthetic DNA"

HYPOTHETICAL: NO

FEATURE:

NAME/KEY: CDS

LOCATION: 9..1238

OTHER INFORMATION: /note="Maize optimized DNA

OTHER INFORMATION: sequence encoding VIP2(a) with the Bacillus secretion signal

OTHER INFORMATION: removed and the eukaryotic secretion signal inserted as

OTHER INFORMATION: contained in pcIB5528"

US-08-463-483A-42

Query Match 2.9%; Score 53.8; DB 2; Length 1241;  
 Best Local Similarity 46.2%; Pred. No. 0.00075;  
 Matches 178; Conservative 0; Mismatches 207; Indels 0; Gaps 0;

Oy 578 ATGAGCATTTGCGCGAGAGAGAGGCGCTCAAGCTTTTGGCATATCCGCTACATGCGAG 637  
 Db 657 AAGAGCATATCAAGCGCCGAGCCCGACAGCTGGGGCATGAAGAACTAGAGAGTGCGCC 716  
 Oy 638 TTTTGCTTCAACAGACAGGTGAGCGGCAAGCTTGAGAGACTCAACTGGGGCTCTTCAAC 697  
 Db 717 AAGGACCTGACCGACAGCGAGCGGCGGCTGAGGCTTACGCGCCGAGAGACTCAAG 776  
 Oy 698 AAGGTGTCAGCGCTGTCTGAGAGAGCGGCGCTACTGATATTCATGCAACTTT 757  
 Db 777 GAGATCAACACTACTGCGGCAACAGGGCGGAGGCGGCAAGAGAGCTGAGCGCCAG 836  
 Oy 758 GCCCGCTACAGCGCGGCGATCGGCCAGGAGGCGGTGCGAGACATCTTTGTGAG 817  
 Db 837 ATCAAGAGATCATGAGCGAGCGCCCTGGGCAAGAGCCCATCCCGAGAACTACACGCTGAC 896  
 Oy 818 CTCTGGGTCAATGCGCAAGTACTAGAGAGAGAGAGAGATCATCTTTGGCGTGAG 877  
 Db 897 CGCTGGTGGGATGCGCGAGTTCGCTAGAGATGAGAGCGCCCTGCGCAAGCTGAG 956  
 Oy 878 AAGAGCGCGAGAGCTGACATTTGAGATCTGGGCGAGAGCTGCCAAAGGTGCTGACT 937  
 Db 957 GACTTCGAGGACAGCTTCTGAAACCATCAAGAGAGAGAGAGGCTATGAGACCCAGC 1016  
 Oy 938 GCGATCCGAAGAGCGCGGCCCACT 962  
 Db 1017 CTGAGCAGCGAGCGCTGCGCCCT 1041

## RESULT 13

US-08-471-046A-39

Sequence 39, Application US/08471046A

Patent No. 5866326

GENERAL INFORMATION:

APPLICANT: Warren, Gregory M

APPLICANT: Kozziel, Michael G

APPLICANT: Mullins, Martha A

APPLICANT: Nye, Gordon J

APPLICANT: Carr, Brian

APPLICANT: Desai, Nalin M

APPLICANT: Kostichka, N. Kristy

APPLICANT: Duck, Nicholas B

TITLE OF INVENTION: Method For Isolating Vegetative Insecticidal

NUMBER OF SEQUENCES: 50

CORRESPONDENCE ADDRESS:

ADDRESSEE: No. 5866326art's Corporation

STREET: 3054 Cornwells Road

CITY: Research Triangle Park

STATE: NC

COUNTRY: USA

ZIP: 27709

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent Release #1.0, Version #1.308

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/471,046A

FILING DATE: 06-JUN-1995

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/463,483

FILING DATE: 05-JUN-1995

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/314,594

FILING DATE: 09-SEP-1994

```

: REGISTRATION NUMBER: 40,403
: REFERENCE/DOCKET NUMBER: CGC 1695/CIP3/DIV6 - SOLV3
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 919-541-8582
: TELEFAX: 919-541-8689
: INFORMATION FOR SEQ ID NO: 42:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1241 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: other nucleic acid
: DESCRIPTION: /desc = "Synthetic DNA"
: FEATURE:
: NAME/KEY: CDS
: LOCATION: 9..1238
: OTHER INFORMATION: /note= "Maize optimized DNA
: OTHER INFORMATION: sequence encoding VIP2A(a) with the Bacillus secretion signal
: OTHER INFORMATION: removed and the eukaryotic secretion signal inserted as
: OTHER INFORMATION: contained in pCIB5528"
US-08-471-044-42

```

```

Query Match      2.9%; Score 53.8; DB 2; Length 1241;
Best Local Similarity 46.2%; Pred. No. 0.00075;
Matches 178; Conservative 0; Mismatches 207; Indels 0; Gaps 0;

```

```

OY 578 ATGAAGATTTCGCCGAAGAGAGAGGCTCAAGCTTTGGCATATCCGCTACATGGCAG 637
DB 657 AAGAGACATCAACGCGGAGGCCACACAGCTGGGGCATGAAGAACTACGAGAGTGGCC 716
OY 638 TTGTCTCAACAACAGCGTGAAGCGGACGAGCTGAGCAGCTCAACTGGGCTCTTACAAC 697
DB 717 AAGAGCTGACCGACGACGAGCGGAGGCGCTGAGCGGCTACGCGCCGACGAGTACAG 776
OY 698 AAGTCTCAACGCTCTCTCGAGAGCGGCGCTACTGATGATTGACAGCACTTT 757
DB 777 GAGATCAACAACCTACCTGCCAACACAGGCGGCGCAACGAGCTGAGCGCCAG 836
OY 758 GCCCGCTACAACGCGGCGCATCATGCGGAGGAGGCGGTGCGGCACTTTGTGAGC 817
DB 837 ATCAAGAACTACAGCGCGGCGCTGGGCAAGAGCCCATCCCGAATACATACCGTGTAC 896
OY 818 CTCTGGTCCAGATCGCAAGATCTACAGAGACAACGACAAGATCATCTTGGCTGATG 877
DB 897 CGCTGGTGGGAGATGCGCGAGTGTGCGCTACCAATACGCGCCCGCCAGCGCTGAG 956
OY 878 AAGGAGCGGACGACCTCGACATTGAGATCTGGCGGACAGAGCTGCCAAGAGTGTGACT 937
DB 957 GACTTCGAGAGAGCGATTCTGAAACACCATCAAGAGAGACAAGGCGCTACATGACACGAC 1016
OY 938 GCGATCCGAAGGCGGCGGCGACT 962
DB 1017 CTGAGCAGCAGCGCGCTGGCGGCT 1041

```

```

RESULT 11
US-08-463-483A-39
: Sequence 39, Application US/08463483A
: Patent No. 5849870
: GENERAL INFORMATION:
: APPLICANT: Warren, Gregory W
: APPLICANT: Kozlowski, Michael G
: APPLICANT: Mullins, Martha A
: APPLICANT: Nye, Gordon J
: APPLICANT: Carr, Brian
: APPLICANT: Desai, Nalin M
: APPLICANT: Kostichka, N. Kristy
: APPLICANT: Duck, Nicholas B
: APPLICANT: Estruch, Juan J
: TITLE OF INVENTION: No. 5849870e1 Pesticidal proteins and strains
: NUMBER OF SEQUENCES: 50
: CORRESPONDENCE ADDRESS:

```

```

: ADDRESSEE: CIBA-GEIGY Corporation
: STREET: 7 Skyline Drive
: CITY: Hawthorne
: STATE: NY
: COUNTRY: USA
: ZIP: 10532
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patent Release #1.0, Version #1.30B
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/463,483A
: FILING DATE:
: CLASSIFICATION: 530
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/314,594
: FILING DATE: 09-SEP-1994
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/218,018
: FILING DATE: 23-MAR-1994
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/037,057
: FILING DATE: 25-MAR-1993
: ATTORNEY/AGENT INFORMATION:
: NAME: Spruill, W. Murray
: REGISTRATION NUMBER: 32,943
: REFERENCE/DOCKET NUMBER: CGC 1695/CIP3
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 919-541-8615
: TELEFAX: 919-541-8615
: INFORMATION FOR SEQ ID NO: 39:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1241 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: other nucleic acid
: DESCRIPTION: /desc = "Synthetic DNA"
: FEATURE:
: NAME/KEY: CDS
: LOCATION: 9..1238
: OTHER INFORMATION: /note= "Maize optimized DNA
: OTHER INFORMATION: sequence encoding VIP2A(a) with the Bacillus secretion signal
: OTHER INFORMATION: removed as contained in pCIB5528"
US-08-463-483A-39

```

```

Query Match      2.9%; Score 53.8; DB 2; Length 1241;
Best Local Similarity 46.2%; Pred. No. 0.00075;
Matches 178; Conservative 0; Mismatches 207; Indels 0; Gaps 0;

```

```

OY 578 ATGAAGATTTCGCCGAAGAGAGAGGCTCAAGCTTTGGCATATCCGCTACATGGCAG 637
DB 657 AAGAGACATCAACGCGGAGGCCACACAGCTGGGGCATGAAGAACTACGAGAGTGGCC 716
OY 638 TTGTCTCAACAACAGCGTGAAGCGGACGAGCTGAGCAGCTCAACTGGGCTCTTACAAC 697
DB 717 AAGAGCTGACCGACGACGAGCGGAGGCGCTGAGCGGCTACGCGCCGACGAGTACAG 776
OY 698 AAGTCTCAACGCTCTCTCGAGAGCGGCGCTACTGATGATTGACAGCACTTT 757
DB 777 GAGATCAACAACCTACCTGCCAACACAGGCGGCGCAACGAGCTGAGCGCCAG 836
OY 758 GCCCGCTACAACGCGGCGCATCATGCGGAGGAGGCGGTGCGGCACTTTGTGAGC 817
DB 837 ATCAAGAACTACAGCGCGGCGCTGGGCAAGAGCCCATCCCGAATACATACCGTGTAC 896
OY 818 CTCTGGTCCAGATCGCAAGATCTACAGAGACAACGACAAGATCATCTTGGCTGATG 877
DB 897 CGCTGGTGGGAGATGCGCGAGTGTGCGCTACCAATACGCGCCCGCCAGCGCTGAG 956
OY 878 AAGGAGCGGACGACCTCGACATTGAGATCTGGCGGACAGAGCTGCCAAGAGTGTGACT 937

```

```

US-08-471-044-39          : Sequence 39, Application US/08471044
Patent No. 5840868
GENERAL INFORMATION:
APPLICANT: Warren, Gregory W
APPLICANT: Koziel, Michael G
APPLICANT: Mullins, Martha A
APPLICANT: Nye, Gordon J
APPLICANT: Carr, Brian
APPLICANT: Desai, Nailal M
APPLICANT: Kostichka, N. Kristy
APPLICANT: Duck, Nicholas B
APPLICANT: Estuch, Juan J
TITLE OF INVENTION: No. 5840868el Pesticidal Proteins and Strains
NUMBER OF SEQUENCES: 50
CORRESPONDENCE ADDRESS:
ADDRESSEE: CIBA-GEIGY Corporation
STREET: 7 Skyline Drive
CITY: Hawthorne
STATE: NY
COUNTRY: USA
ZIP: 10532
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30B
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/471,044
FILING DATE: 06-JUN-1995
CLASSIFICATION: 800
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/463,483
FILING DATE: 05-JUN-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/314,594
FILING DATE: 09-SEP-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/218,018
FILING DATE: 23-MAR-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/037,057
FILING DATE: 25-MAR-1993
ATTORNEY/AGENT INFORMATION:
NAME: Pace, Gary M.
REGISTRATION NUMBER: 40,403
REFERENCE/DOCKET NUMBER: CGC 1695/CIP3/DIV6 - SOLV3
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-541-8582
TELEFAX: 919-541-8689
INFORMATION FOR SEQ ID NO: 39:
SEQUENCE CHARACTERISTICS:
LENGTH: 1241 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: other nucleic acid
DESCRIPTION: /desc = "Synthetic DNA"
HYPOTHETICAL: NO
FEATURE:
NAME/KEY: CDS
LOCATION: 9..1238
OTHER INFORMATION: /note= "Maize optimized DNA
OTHER INFORMATION: sequence encoding VIP2A(a) with the Bacillus secretion signal"
OTHER INFORMATION: removed as contained in pcib5527"
US-08-471-044-39
Query Match      2.9%; Score 53.8; DB 2; Length 1241;
Best Local Similarity 46.2%; Pred. No. 0.00075;
Matches 178; Conservative 0; Mismatches 207; Indels 0; Gaps 0;

```

```

Db 657 AAGAAAGCATCAACAGCCCGAGGCGCCACAGCGTGGGGCATGATAGAACTACAGAGAGATGGGCC 716
Oy 638 TTGTGCTCTACAACAACGCGTGGACGGCCAGCTGGAGCACTCACTGGGGCTCTTCAACA 697
Db 717 AAGGACCTGACCGCACACCCGAGCCCGAGGCCCTGGACGGCTACGGCCCGCAGGACTTCAAG 776
Oy 698 AAGGTGCTCAACGGCTGTCTCTGAGAGGGGGGCCCTACTGCGATGATTGACATGCACAACTTT 757
Db 777 GAGATCAACAACATCACTGCGCCCAACCAAGGGGGGAGCGGACGACAGAACTGGAGCCCGAG 836
Oy 758 GCCCGCTACAACGCGCGGCATCATCGGCCCAAGGGAGCGCTGTGCGAGACATCTTTGTGCA 817
Db 837 ATCAAGAAACATCAAGCGACGCGCCCTGGGGCAAGAACCCATCCCGAGAAATCAACCTGAC 896
Oy 818 CTCTGGGCTCAGATCGCAAGTACTACAGAGGACACGACAGATCATCTTTGGCTGATG 877
Db 897 CGCTGTGTCGGCATGCCCGGATTCGGCTCTACAGATCAGCGACCCCTGCGCAGCCTGGA 956
Oy 878 AAGCAGCGCGACGACCTCTGCACTTGAAGATCTGGGGCGCAGCGCAAGAGTGTGCTACT 937
Db 957 GACTTCGAGGAGCGAGCTTCTGGAACACCATCAAGAGGAGACAGAGGCTTACTGACACAC 1016
Oy 938 GCGATCCGAAAGCGCGCGCCACT 962
Db 1017 CTGAGCAGCGAGCGCCCTGGCGGCT 1041

RESULT 10
US-08-471-044-42
Sequence 42, Application US/08471044
Patent No. 5840868

GENERAL INFORMATION:
APPLICANT: Warren, Gregory M
APPLICANT: Koziele, Michael G
APPLICANT: Mullins, Martha A
APPLICANT: Nye, Gordon J
APPLICANT: Carr, Brian
APPLICANT: Desai, Nalin M
APPLICANT: Koslicka, N. Kristy
APPLICANT: Duck, Nicholas B
APPLICANT: Estruch, Juan J
TITLE OF INVENTION: No. 5840868el Pesticidal Proteins and Strains
NUMBER OF SEQUENCES: 50
CORRESPONDENCE ADDRESS:
ADDRESSEE: CIBA-GEIGY Corporation
STREET: 7 Skyline Drive
CITY: Hawthorne
STATE: NY
COUNTRY: USA
ZIP: 10532

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30B
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/471,044
FILING DATE: 06-JUN-1995
CLASSIFICATION: 800
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/463,483
FILING DATE: 05-JUN-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/314,594
FILING DATE: 09-SEP-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/218,018
FILING DATE: 23-MAR-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/037,057
FILING DATE: 25-MAR-1993
ATTORNEY/AGENT INFORMATION:
NAME: Pace, Gary M.

```

```

STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: other nucleic acid
DESCRIPTION: /desc = "Synthetic DNA"
HYPOTHETICAL: NO
FEATURE:
NAME/KEY: CDS
LOCATION: 9..1238
OTHER INFORMATION: /note= "Maize optimized DNA
OTHER INFORMATION: sequence encoding VIP2(a) with the Bacillus secretion signal"
OTHER INFORMATION: removed as contained in pc1B5527"
US-08-471-033-39

```

```

Query Match      2.9%; Score 53.8; DB 1: Length 1241;
Best Local Similarity 46.2%; Pred. No. 0.00075;
Matches 178; Conservative 0; Mismatches 207; Indels 0; Gaps 0;

```

```

OY 578 ATGAAGCATTTCCGCGAAGACGAGCGGCTCAAGCTTTTGGCATATCGGCTACATGGCAG 637
Db 657 AAGAACGACATCAACGCGGAGGCCACAGCTGGGGCATGAAGAACTACGAGAGTGGGCC 716
OY 638 TTGTCTCTCAACAACAGCGTGGAGCGCAAGCTGGAGAGCTCAACTGGGCTCCTTACAAC 697
Db 717 AAGGACCTGACCGACGACGCGGAGGCCCTGGACGGCTACGCGCCCGCAGACTACAG 776
OY 698 AAGTCTCAACGCGCTCTCTGAGAGCGGCGGCTACTGATGATTTGACATGCACAACTTT 757
Db 777 GAGATCAACAACATCTCTGCGCAACGAGGCGGCGGCAACGAAAGCTGAGCGCCAG 836
OY 758 GCCCGCTACAAGCGGCGCATCATGCGGCGGAGGAGCGGCTGCGGCAATCTTTGTGCGAC 817
Db 837 ATCAAGAACATCAACGCGCGGCTGGGCAAGAGCCCATCCCGAAGCATATCAGCTGTAC 896
OY 818 CTCTGGTCTCAAGATCGCAAGATCTACGAGACGACGACAAAGATCATTTTGGCTGATG 877
Db 897 CGGTGGTGGGCGCATGCGCGAGTTGGGTACCATGACGACGCCCTGCCAGGCTGAG 956
OY 878 AAGGACCGGACGACGCTCTGACATTTGAGATCTGGCGGCGAGAGCTGCCAAGAGTGTGCT 937
Db 957 GACTTCGAGAGGAGCGATTCGTAACACCATCATGAGAGGAGCAAGGCGCTACATGAGCAGC 1016
OY 938 GCGATCCGAAGGCGGCGGCGACCT 962
Db 1017 CTGAGCAGCAGCGGCTGGCGGCT 1041

```

## RESULT 8

```

US-08-471-033-42
Sequence 42, Application US/08471033
Patent No. 5770696

```

## GENERAL INFORMATION:

```

APPLICANT: Warren, Gregory M
APPLICANT: Koziele, Michael G
APPLICANT: Mullins, Martha A
APPLICANT: Nye, Gordon J
APPLICANT: Carr, Brian
APPLICANT: Desai, Nalin M
APPLICANT: Kostichka, N. Kristy
APPLICANT: Duck, Nicholas B
APPLICANT: Estruch, Juan J
TITLE OF INVENTION: No. 5770696el Pesticidal Proteins and Strains
NUMBER OF SEQUENCES: 50
CORRESPONDENCE ADDRESS:
ADDRESSEE: CIBA-GEIGY Corporation
STREET: 7 Skyline Drive
CITY: Hawthorne
STATE: NY
COUNTRY: USA
ZIP: 10532

```

```

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS

```

```

SOFTWARE: PatentIn Release #1.0, Version #1.30B
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/471,033
FILING DATE:
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/314,594
FILING DATE: 09-SEP-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/218,018
FILING DATE: 23-MAR-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/037,057
FILING DATE: 25-MAR-1993
ATTORNEY/AGENT INFORMATION:
NAME: Pace, Gary M.
REGISTRATION NUMBER: P-40,403
REFERENCE/DOCKET NUMBER: CGC 1695/CIP3/DIV7 - SOLV3
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-541-8582
TELEFAX: 919-541-8689
INFORMATION FOR SEQ ID NO: 42:
SEQUENCE CHARACTERISTICS:
LENGTH: 1241 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: other nucleic acid
DESCRIPTION: /desc = "Synthetic DNA"
HYPOTHETICAL: NO
FEATURE:
NAME/KEY: CDS
LOCATION: 9..1238
OTHER INFORMATION: /note= "Maize optimized DNA
OTHER INFORMATION: sequence encoding VIP2(a) with the Bacillus secretion signal
OTHER INFORMATION: removed and the eukaryotic secretion signal inserted as
OTHER INFORMATION: contained in pc1B5528"
US-08-471-033-42

```

```

Query Match      2.9%; Score 53.8; DB 1: Length 1241;
Best Local Similarity 46.2%; Pred. No. 0.00075;
Matches 178; Conservative 0; Mismatches 207; Indels 0; Gaps 0;

```

```

OY 578 ATGAAGCATTTCCGCGAAGACGAGCGGCTCAAGCTTTTGGCATATCGGCTACATGGCAG 637
Db 657 AAGAACGACATCAACGCGGAGGCCACAGCTGGGGCATGAAGAACTACGAGAGTGGGCC 716
OY 638 TTGTCTCTCAACAACAGCGTGGAGCGCAAGCTGGAGAGCTCAACTGGGCTCCTTACAAC 697
Db 717 AAGGACCTGACCGACGACGCGGAGGCCCTGGACGGCTACGCGCCCGCAGACTACAG 776
OY 698 AAGTCTCAACGCGCTCTCTGAGAGCGGCGGCTACTGATGATTTGACATGCACAACTTT 757
Db 777 GAGATCAACAACATCTCTGCGCAACGAGGCGGCGGCAAGAGCCCATCCCGAAGCATATCAGCTGTAC 896
OY 758 GCCCGCTACAAGCGGCGCATCATGCGGCGGAGGAGCGGCTGCGGAGCATTTTGTGCGAC 817
Db 837 ATCAAGAACATCAACGCGCGGCTGGGCAAGAGCCCATCCCGAAGCATATCAGCTGTAC 896
OY 818 CTCTGGTCTCAAGATCGCAAGATCTACGAGACGACGACAAAGATCATTTTGGCTGATG 877
Db 897 CGGTGGTGGGCGCATGCGCGAGTTGGGTACCATGACGACGCCCTGCCAGGCTGAG 956
OY 878 AAGGACCGGACGACGCTCTGACATTTGAGATCTGGCGGCGAGAGCTGCCAAGAGTGTGCT 937
Db 957 GACTTCGAGAGGAGCGATTCGTAACACCATCATGAGAGGAGCAAGGCGCTACATGAGCAGC 1016
OY 938 GCGATCCGAAGGCGGCGGCGACCT 962
Db 1017 CTGAGCAGCAGCGGCTGGCGGCT 1041

```

## RESULT 9



```

ADDRESS: Genencor International
STREET: 925 Page Mill Road
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94080
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/382,452D
FILING DATE: February 1, 1995
ATTORNEY/AGENT INFORMATION:
NAME: Christopher L. Stone
REGISTRATION NUMBER: 36,696
REFERENCE/DOCKET NUMBER: GC226-2
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 742-7555
TELEFAX: (415) 742-7217
INFORMATION FOR SEQ ID NO: 15:
SEQUENCE CHARACTERISTICS:
LENGTH: 1155 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: join(1..56, 231..1158)
US-08-382-452D-15

Query Match
Best Local Similarity 15.1%; Score 275.6; DB 4; Length 1155;
Matches 489; Conservative 0; Mismatches 334; Indels 3; Gaps 1;

```

```

QY 1103 GACATCAACAACCTCGGCTGCACGCCGAGTGCACACAGACAC--GTGAGCCCTTC 1159
DB 838 GACTCAGACAACCTCGGCTGCACGCCGAGTGCACACAAATGATGACGGCCCTT 897
QY 1160 AAGCACTTCGGGACTGGCTGAGCAGAAAGGCCCGCCATCATCTCCGAAGCGGC 1219
DB 898 TCTCCGCTTGGCCACTTGGCTGCAGAAACAAATGCGCAGGATATCTGCAAGAACCGGT 957
QY 1220 GCGTCATGAAACCTTCGTGCATGACTGCTTCCGCCGCCAGAACAGCCATTAGCGAA 1279
DB 958 GGTGCACACGTTCACTGCTGCATACAGACATGTGCCAGCAATTCATATCTCAACCG 1017
QY 1280 AACAGCGAGTCTACATTGCTTGTGGCTGGGCTGGCCGACCTTTGACACGTCGTAC 1339
DB 1018 AACTCAGATGCTTACTGCTATGTTGTTGGGTGGCGGATCATTTGATAGACGATAT 1077
QY 1340 ATCTGATCTGACTCCCTCGGCAAGCCCGGCACTACACCGACA 1385
DB 1078 GTCTGACGAAACCGACTAGCAGTGTACTGATGACGAGACA 1123

```

```

RESULT 5
US-08-232-463-14
Sequence 14, Application US/08232463
Patent No. 5670367
GENERAL INFORMATION:
APPLICANT: DORNER, F.
APPLICANT: SCHEIFLINGER, F.
APPLICANT: FALKNER, F. G.
TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
NUMBER OF SEQUENCES: 52
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Foley & Lardner
STREET: 1800 Diagonal Road, Suite 500
CITY: Alexandria
STATE: VA
COUNTRY: USA
ZIP: 22313-0299
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/232,463
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/07/935,313
FILING DATE:
APPLICATION NUMBER: EP 91 114 300.6
FILING DATE: 26-AUG-1991
ATTORNEY/AGENT INFORMATION:
NAME: BENT, Stephen A.
REGISTRATION NUMBER: 29,768
REFERENCE/DOCKET NUMBER: 30472/114 IMMU
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 836-9300
TELEFAX: (703) 883-4109
TELEX: 899149
INFORMATION FOR SEQ ID NO: 14:
SEQUENCE CHARACTERISTICS:
LENGTH: 7218 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
CLONE: pTZ9pl-P15
US-08-232-463-14

Query Match
Best Local Similarity 5.2%; Score 94.2; DB 1; Length 7218;
Matches 6; Conservative 259; Mismatches 112; Indels 0; Gaps 0;

```

Query Match	Score	DB 2;	Length
Best Local Similarity	59.28;	Pred. No. 3.1e-57;	

CORRESPONDENCE ADDRESS:

CORRESPONDENCE ADDRESS:

```

QY 623 TCCGCTCATGGCAGTTTGTCTTCAACAACAGCGTGGACGGCAAGCTGACGAGCTCAAC 682
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 653 CCCGTCGAGTGGCAGTACCTCGTAAACAAATCGGGTGGAACTCTGATTCACCAAT 712
QY 683 TGGGGCTCCACAAACAGGTGCTCAAGCCGTGTGAGAGGGCGGCTCTGATCATATT 742
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 713 ATCTCGAAGTATGATGCTGCTGTCAGGGGGGCTGTCTCTGCTGATATCTGATATC 772
QY 743 GACATGCAACAATTTGCCCCCTACAAAGCGCGCATTCGCGCAGGAGCGGTTCGGAC 802
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 773 GACATCCACAATATATGTCATGAGAGGTGATCATTTGGCCAGGAGCGCCCTACAAAT 832
QY 803 GACATCTTTGTCAGCTCTGGGTCCAGATGCAAAATGACTACAGAGCAACAGCAAGATC 862
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 833 GCCCAGATTTACCAATCTTGTGTCGCACTTGGCAAGTACGAGCTCTAGTCAAGGATG 892
QY 863 ATCTTTGGCCGTGATGAGACGGCGGACGCTGACATTCAGATCTGGCGGACAGCTGC 922
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 893 TGGTTGGAATTAATGAAATGAGCCCGACGAGCTGAACATCAACACTTGGGCTGCCAGGTT 952
QY 923 CAAAAGCTGCTCACTGCGATCGGAAGCGCGCGCACCTGCGAGATGATCTCTGCTGCC 982
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 953 CAAGAGGTGCTCACTGCAATCCGGAACGGCGGTGCTAGTGCAATGATTTCTCTGCTCT 1012
QY 983 GGAACCAACTTTGCCAGCTGCGAGACGTATGTCTCACTGGCAGCGCGGAAGCCCTGGC 1042
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1013 GGAATATATATATCAATCTGCGGCGAGCTTTATTTCCGATGGCAGTCCAGCCCTGTCT 1072
QY 1043 AGATTTAGAACCCGGATGGAGACCGGATTTGCTGATCTTGTATGTCACAGTATCTC 1102
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1073 GAGGTAAACGAACTCTGATGATCAACAACAGATTAATCTTCGATGTCACAGTACTTA 1132
QY 1103 GACATCAACAACCTCGGGGTGCGACCGCGAGTGCACACAGACAACGTCGA---GCGCTTC 1159
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1133 GACTCGGACAACTCGGCTACTCAAGCGGAATGCACTACAAACAACATCGAGCGGCTTT 1192
QY 1160 AAGCACTTCCGGAGCTGCGTGAAGCAGAAAGAGCGGCAATCATCTCCGAAGCGGC 1219
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1193 GCTCCCTGCTCCACTGCTTGCAGAGCAACCGCAGGCTATTTGAGAGGAAACCGGC 1252
QY 1220 GCGTCAATGGAACCTTGTGATGATGATGCTTGTGCGCCAGAAACAGAGCATTAAGCAA 1279
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1253 GGTGGCAATTTTCAGTCTCTGATCCAGATTTGTGCCAACAAGATCCAGTCTCAACAG 1312
QY 1280 AACAGGACCTCTACATTTGGCTTTGCGGTGGGGTGGCGGACCTTTGACAGCTCTAC 1339
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1313 AACTCAAGATGTCTATCTTGGCTATGCTGCTGGGGTCCGCTTCAATTTGATAGCACTTAT 1372
QY 1340 ATCTTGACTGTGATCCCTCCGCGCAAGCCCGGCACTACACCGCAACAAGCTCATAC 1399
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1373 ATTCTGACGGAAGCGCTACTGGAAGGGTAACTGCTGAGCGACATCTCCCTAGTTAGC 1432
QY 1400 GAGTG 1404
    |||
Db 1433 TCGTG 1437

```

```

: CITY: South San Francisco
: STATE: CA
: COUNTRY: USA
: ZIP: 94080
:
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patent In Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/169,948B
: FILING DATE: DEC 17 1993
: CLASSIFICATION: 435
: ATTORNEY/AGENT INFORMATION:
: NAME: Horn, Margaret A.
: REGISTRATION NUMBER: 33,401
: REFERENCE/DOCKET NUMBER: GC226
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (415) 742-7536
: TELEFAX: (415) 742-7217
: INFORMATION FOR SEO ID NO: 15
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1155 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: DNA (genomic)
: FEATURE:
: NAME/KEY: CDS
: LOCATION: join(1..56, 231..1155)
:
: US-08-169-948B-15
:
: Query Match 15.1%; Score 275.6; DB 2; Length 1155;
: Best Local Similarity 59.2%; Pred. No. 3.1e-57;
: Matches 489; Conservative 0; Mismatches 334; Indels 3; Gaps 1;
:
QY 563 GATGGCCCGCGCCAGATGAAGCATTTGCCGAGAGACAGCGGCTCAACGCTTTGGCATA 622
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 298 GATGGCATCTGGCGAGATGACGACTTCTCAACGAGAGCGGATGACTATTTCCGCTTA 357
QY 623 TCCGCTCATGGCAGTTTGTCTTCAACAACAGCGTGGACGGCAAGCTGACGAGCTCAAC 682
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 358 CTTGCGATGCGACTGCTCTGTCACACAAATTTGGCGGCAATCTTATTCACAGAC 417
QY 683 TGGGGCTCTACAAACAAGGCTGTCACGCGCTGTCTCGAGAGGGCGGCTACTGATATT 742
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 418 ATTCCGAAGTATGATGACGCTTGTTCAGGGGTGCTGTCTGGCGGATCTGATGATGTC 477
QY 743 GACATGCAACAATTTGCCCGCTACACGCGGCGATCATTCGCGCAGGAGCGGTGCGAC 802
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 478 GACATCCACAATTAATGCTGATGGAACGGTGGATCATTTGCTCAGGGCGGCTTACTAAT 537
QY 803 GACATCTTTGTCAGACCTCTGGGTCCAGATGCGCAAGTACTACGAGGACAAACAGCAATC 862
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 538 GCTCAATTCAGAGAGCTTTGTGTCAGATGGCATCAAGTACGATCTCACTGAGGCTG 597
QY 863 ATCTTTGGCCGTGATGAAGAGCGCGACGACCTGACATTTGAGATCTGGGCGAGAGCTGC 922
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 598 TGGTTGCGCATGATGATGAGAGCCCGACGAGGTGAATCAACACCTGGGCTGCCAGGTC 657
QY 923 CAAAAGTGTGCTACGTGCGATCCGAAGCGCGCGCACCTGCGAGATGATCTCTGCC 982
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 658 CAAGAGGTGTAACCGCAATCCGCAACGCTGTGCTACGTGCGCAATTCATCTTGTCC 717
QY 983 GGAACCAACTTTGGCAGCTGAGAGAGCTATGTGTCACCTGCGAGCGGAGGAGCCCTGGG 1042
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 718 GGAATATATTTGGCATCTGCTGGGGCTTTTATATCCATGCACTGACGCGCCCTGTCT 777
QY 1043 AAGATTAAGAACCCGAGTGAAGACGAGATTTGCTGACTTTGATGTCCAAAGTATCTC 1102
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 778 CAAGTCAACGAAACCGGATGGGTAAACAAGCAATCTGATTTTGTGACGTCACAAATACTTG 837
QY 1103 GACATCAACAACCTCGGGGTGCGACGCGGAGTGCACCAACAGCAAC---GTGAGCGCTTC 1159

```

GenCore version 5.1.4.p5.4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 15, 2003, 02:49:23 ; Search time 70.2978 Seconds  
(without alignments)  
7965.992 Million cell updates/sec

Title: US-10-028-245-1

Perfect score: 1826  
Sequence: 1 gtcagccacgcgtccgtc.....aatcaaaaaaaaaaaaaa 1826

Scoring table: IDENTITY\_NUC  
Gapop 10.0, Gapext 1.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 08

Maximum Match 1008  
Listing first 45 summaries

Database :

Issued\_Patents\_NA:\*  
1: /cgn2\_6/prodata/1/lna/5A.COMB.seq:\*  
2: /cgn2\_6/prodata/1/lna/5B.COMB.seq:\*  
3: /cgn2\_6/prodata/1/lna/6A.COMB.seq:\*  
4: /cgn2\_6/prodata/1/lna/6B.COMB.seq:\*  
5: /cgn2\_6/prodata/1/lna/PCTUS.COMB.seq:\*  
6: /cgn2\_6/prodata/1/lna/Packlist1.seq:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	289.8	15.9	1463	4	US-09-254-733-6
2	275.6	15.1	1155	2	US-08-169-9488-15
3	275.6	15.1	1155	2	US-08-448-873-15
4	275.6	15.1	1155	4	US-08-382-452D-15
5	94.2	5.2	7218	1	US-08-232-463-14
6	55.6	3.0	390	4	US-09-197-649-7
7	53.8	2.9	1241	1	US-08-471-033-39
8	53.8	2.9	1241	1	US-08-471-033-42
9	53.8	2.9	1241	2	US-08-471-044-39
10	53.8	2.9	1241	2	US-08-471-044-42
11	53.8	2.9	1241	2	US-08-463-483A-39
12	53.8	2.9	1241	2	US-08-463-483A-42
13	53.8	2.9	1241	2	US-08-471-046A-39
14	53.8	2.9	1241	2	US-08-471-046A-42
15	53.8	2.9	1241	2	US-08-470-566B-39
16	53.8	2.9	1241	2	US-08-470-566B-42
17	53.8	2.9	1241	2	US-08-469-334-39
18	53.8	2.9	1241	2	US-08-469-334-42
19	53.8	2.9	1241	3	US-09-300-529-39
20	53.8	2.9	1241	3	US-09-300-529-42
21	53.8	2.9	1358	1	US-08-471-033-45
22	53.8	2.9	1358	2	US-08-471-044-45
23	53.8	2.9	1358	2	US-08-463-483A-45
24	53.8	2.9	1358	2	US-08-471-046A-45
25	53.8	2.9	1358	2	US-08-470-566B-45
26	53.8	2.9	1358	2	US-08-469-334-45
27	53.8	2.9	1358	3	US-09-300-529-45

28	53.8	2.9	1389	1	US-08-471-033-27	Sequence 27, Appl
29	53.8	2.9	1389	2	US-08-471-044-27	Sequence 27, Appl
30	53.8	2.9	1389	2	US-08-463-483A-27	Sequence 27, Appl
31	53.8	2.9	1389	2	US-08-471-046A-27	Sequence 27, Appl
32	53.8	2.9	1389	2	US-08-470-566B-27	Sequence 27, Appl
33	53.8	2.9	1389	2	US-08-469-334-27	Sequence 27, Appl
34	53.8	2.9	1389	3	US-09-300-529-27	Sequence 27, Appl
35	53.8	2.9	1389	1	US-08-471-033-24	Sequence 24, Appl
36	53.8	2.9	1389	2	US-08-471-044-24	Sequence 24, Appl
37	53.8	2.9	1389	2	US-08-463-483A-24	Sequence 24, Appl
38	53.8	2.9	1389	2	US-08-471-046A-24	Sequence 24, Appl
39	53.8	2.9	1389	2	US-08-470-566B-24	Sequence 24, Appl
40	53.8	2.9	1389	2	US-08-469-334-24	Sequence 24, Appl
41	53.8	2.9	1389	3	US-09-300-529-24	Sequence 24, Appl
42	53.8	2.9	4031	1	US-08-471-033-49	Sequence 49, Appl
43	53.8	2.9	4031	2	US-08-471-044-49	Sequence 49, Appl
44	53.8	2.9	4031	2	US-08-463-483A-49	Sequence 49, Appl
45	53.8	2.9	4031	2	US-08-471-046A-49	Sequence 49, Appl

#### ALIGNMENTS

RESULT 1  
US-09-254-733-6

Sequence 6, Application US/09254733

Patent No. 6277596

GENERAL INFORMATION:

APPLICANT: WATANABE, MANABU

APPLICANT: MORIYA, TATSUKI

APPLICANT: AOYAGI, KAORI

APPLICANT: SUMIDA, NAOMI

APPLICANT: MURAKAMI, TAKESHI

TITLE OF INVENTION: REGULATORY SEQUENCE OF CELLULOSE CBH1 GENES ORIGINATING

TITLE OF INVENTION: IN TRICHODERMA VIRIDE AND SYSTEM FOR MASS-PRODUCING

FILE REFERENCE: 99-0266\*/LC(WMC)/00144

CURRENT APPLICATION NUMBER: US/09/254,733

CURRENT FILING DATE: 1999-05-07

NUMBER OF SEQ ID NOS: 52

SOFTWARE: PatentIn Ver. 2.0

SEQ ID NO 6

LENGTH: 1463

TYPE: DNA

ORGANISM: TRICHODERMA VIRIDE MC300-1

FEATURE:

NAME/KEY: sig\_peptide

LOCATION: (14)..(76)

FEATURE:

NAME/KEY: CDS

LOCATION: (14)..(76)

FEATURE:

NAME/KEY: mat\_peptide

LOCATION: (77)..(1450)

FEATURE:

NAME/KEY: CDS

LOCATION: (77)..(342)

FEATURE:

NAME/KEY: Intron

LOCATION: (343)..(525)

FEATURE:

NAME/KEY: CDS

LOCATION: (526)..(1450)

FEATURE:

NAME/KEY: CDS

LOCATION: (526)..(1450)

FEATURE:

NAME/KEY: CDS

LOCATION: (526)..(1450)

FEATURE:

NAME/KEY: CDS

LOCATION: (526)..(1450)

Query Match 15.9%; Score 289.8; DB 4; Length 1463;  
Best Local Similarity 59.8%; Pred. No. 1.3e+60;  
Matches 505; Conservative 0; Mismatches 337; Indels 3; Gaps 1;

QY 563 GATGGCCGCGCCAGATGAGCATTTTCGCCGGAAGACGAGCGGCTCAACGCTTTCGCATA 622  
DB 593 GACGGTATGCGCCAGATGAGCATTTTCGCCGGAAGACGAGCGGCTCAACGCTTTCGCATA 652

**THIS PAGE BLANK (USPTO)**



QY 884 CCGCAGCAGCTCGACATTTGAGATCTGGGCGCAGACGTGCCAAGAGTCTGCTACTGCCATC 943.  
 DB 487 TACACGACATGACGAAAGGCTTGTCTCCAGCTCAACGAGCGCCATCGACGAGATC 546  
 QY 944 CGAAGGCGGGCGCCACCTGCGAGATATCTCTGCGCCGGAACCACTTTGCCAGGTC 1003  
 DB 547 CGCGCCCGGGCGCCACATCATCACTTACATCTTCTCGAGGGCACTCTGAGCGGGGCC 606  
 QY 1004 GAGACGATGTGTCATCGAGCGGGAAGCCCTCGGCAAGATTACGAACCGCGATGA 1063  
 DB 607 TGGACAT-----GACCCAGGTCAACACGAGGATGGGAACCTTACGAC 651  
 QY 1064 AGCAGCAGATTTGCTGATCTTGTATGTCACAAATATCTGACATCAACATCCGGGTG 1123  
 DB 652 CCGCAGAAACAAAGATCGTGTAGAGATGACAGATCTGAGCGGAGCGGTGGGACG 711  
 QY 1124 CAGCCGAGTGCACACAGACAGACGTC-----GACGCTTCAACGACTTGGCGGACTG 1177  
 DB 712 TCGGACGAGTGCCTCAACCTGACCATCGGGCAGACCGCGTGAAGTGGCGGCGCTGG 771  
 QY 1178 CTGAGGAGCAAGACAGCGCCAGCCATCATCTCGAAGGCGCGTCCATGAACTTTCG 1237  
 DB 772 CTGAGAGAGAAAGCGCAAGAGGCGATCTGGCGCAGTACGCTGGCGGCGCAACAGCGTG 831  
 QY 1238 TGCATGACTGCTTCTGCGCCAGACAAAGGCGCATTAAGCGAAACAGCGAGCTTACAT 1297  
 DB 832 TGCAGAGCGCGCGTACCGGCGATCTGACATCTCGCCCAACATACGATGTCTGAC 891  
 QY 1298 GCGTTTGGGCTGGGGTGGCGG 1320  
 DB 892 GGTCTATCTGTGTGGCGGCTGG 914

RESULT 14

ID ABQ44816/c  
 AC ABQ44816; standard; DNA; 985 BP.

DT 12-JUL-2002 (first entry)

XX Oligonucleotide for detecting cytosine methylation SEQ ID NO 31407.

KM Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;  
 KW drug; side effect; cancer; central nervous system; cardiovascular;  
 KM gastrointestinal; respiratory system; single nucleotide polymorphism;  
 KW SNP; cell differentiation; ds.

XX Homo sapiens.

OS WO200218632-A2.

PN 07-MAR-2002.

PD 01-SEP-2001; 2001MO-EP10074.

PF 01-SEP-2000; 2000DE-1043826.

PR 05-SEP-2000; 2000DE-1044543.

XX (EPIC-) EPIGENOMICS AG.

PI Olek A, Piepenbrock C, Berlin K, Guelig D;

DR WPI; 2002-371829/40.

PT Determining the degree of cytosine methylation in genomic DNA, useful  
 for diagnosis and prognosis, comprises selective hybridization of  
 PT amplicons from chemically treated DNA

PS Claim 12: 56pp + Sequence Listing; 56pp: German.

CC This invention describes a novel method for determining the degree of  
 CC methylation of a particular cytosine in a motif 5'-CpG-3', present in a

CC genomic sample of DNA. The sample is treated chemically to convert  
 CC cytosine (C) but not methylated C, to uracil, then part of the genomic  
 CC DNA that contains the target C is amplified to form a labeled amplicon.  
 CC The amplicon is hybridised to two classes, each with at least one  
 CC member, of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers  
 CC and the degree of hybridisation to both classes is determined from the  
 CC label on the amplicon. From the ratio of labels hybridised to the two  
 CC classes of oligomers, the degree of methylation is calculated. The method  
 CC is used: (1) for diagnosis and/or prognosis of side effects of  
 CC therapeutic drugs and of a wide range of diseases, e.g. cancer, disorders  
 CC of the central nervous, cardiovascular, gastrointestinal and respiratory  
 CC systems etc., particularly by detecting mutations or single nucleotide  
 CC polymorphisms (SNP's); and (11) for differentiation of cell or tissue  
 CC types and for investigating cell differentiation. The method allows the  
 CC methylation status of many C residues to be determined simultaneously.  
 CC AB013410-AB054121 represent genomic DNA sequences used to illustrate the  
 CC method for determining the degree of cytosine methylation described in  
 CC the disclosure of the invention.

XX SQ Sequence 985 BP; 234 A; 145 C; 326 G; 280 T; 0 other;

Query Match 4.1%; Score 74.2; DB 24; Length 985;  
 Best Local Similarity 46.7%; Pred. No. 2.7e-08; ~ Mismatches 8; Gaps 3;

Matches 342; Conservative 0; Mismatches 383; Indels 8; Gaps 3;

QY 442 CCTCGCCGCGCAAGATCAATATCTGGGCGTCCCATTTCCGGAATGCACTTTGGCTCGA 501  
 DB 840 CGTGCAGGTGACATCGACATCGACATCGACATCGACATCGACATCGACATCGACATCGA 781  
 QY 502 CATGACGCGGACGTGCTGCGACTGACAGTGTGTCGCTGCTGCTGCTGCTGCTGCTGCTG 561  
 DB 780 CGTCTACATCGAC--GTGACATCGACATCGACATCGACATCGACATCGACATCGACAT 723  
 QY 562 AGATGGGCGCGCGGCGGAGATGGAATTTGCGGAGAGAGAGCGGCGCTCAACGCTTTCGAT 621  
 DB 722 CTACATGACGCTCAACGTCACATCTACATGACATCAACATCTACATGACATCTACAT 663  
 QY 622 ATCCGCTACATGAGCAGTTTGTCTCAACACACGCTGAGCGCAAGCTGAGCGAGCTCAA 681  
 DB 662 CGACATGAGCGTGCACATCGACATCGACATCGGCTGTACATCTACATCTACATCTACAT 603  
 QY 682 CTGGGCGCTCTACAAACAGGTCTCAACGCTGTCTGAGAGCGGCGCTTACTGATGAT 741  
 DB 602 CTACATGACGCTCGACATCTACATCTACATCTACATCTACATCTACATCTACATCTAC 543  
 QY 742 TGACATGACAACTTTGGCGGCTCAACGCGGCGGATCATGCGGCGGAGGCGGTGCGA 801  
 DB 542 CGACGCTACATCTACATCTACATCTACATCTACATCTACATCTACATCTACATCTACAT 483  
 QY 802 CGACATCTTTGTCGACCTCTGCTGCTCAGAT--CGCAAAGTACTACGAGGACGACAA 858  
 DB 482 CTACATCTACATCTACATCTACATCTACATCTACATCTACATCTACATCTACATCTAC 423  
 QY 859 GATCATCTTTGGCGCTGATGAGAGCGGCGGACGCTGACATTTGATCTGCGGCGAGC 918  
 DB 422 CTACATCTACATCTACATCTACATCTACATCTACATCTACATCTACATCTACATCTAC 363  
 QY 919 GTGCAAAAGGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 978  
 DB 362 CTACGTCGACATCGACATCGACATCGACATCGACATCGACATCGACATCGACATCGAT 303  
 QY 979 GCCCGGAACCAATTTGCGACGCTGAGAGCTATGTCATCGACGCGGAGAGCCCT 1038  
 DB 302 CTACGTCGACATCTACATCTACATCTACATCTACATCTACATCTACATCTACATCTAC 243  
 QY 1039 CGCGAAGATTACGACCCGGA---TGAAGCAGCGGATTTGCTGATCTTGTGATGTCACAA 1095  
 DB 242 CGACATCGACATCAACGCTGACATCGACATCAACATCTACATCTACATCTACATCTACAT 183  
 QY 1096 GTATCTGACATCAACACTCGGCTGCGACGCGGAGTGCACACAGACAGCTGACGC 1155  
 DB 182 CTACGTCGACATCAACATCAACATCTACATCTACATCTACATCTACATCTACATCTACAT 123

Query Match 4.3%; Score 79.4; DB 19; Length 912;  
 Best Local Similarity 47.5%; Pred. No. 1.4e-09;  
 Matches 365; Conservative 0; Mismatches 371; Indels 33; Gaps 3;

596 GAGGAGGGCTCAAGCTTTTCCGATATCCGATACATGGAGTTTGTCTCAACACACG 655  
 121 GACAAAGGGGATGAATCTTCCCGCTCCAGTTTATGATGAGAGGTTGTGCGCCGATCG 180  
 656 GTGGAGCGCAAGCTGAGACGCTCAACTGGGGCTCCCTACAAAGGTCGTCGCTGT 715  
 181 ATGACTGCTTCAATATGATGAGATCTGTGCGCACTTGACGACAGTATGATTAAGCGGTA 240  
 716 CTGAGACGGGCGCTACTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 775  
 241 ACGAGCGAGAGCGCCATGGCTTGTGACCTTCATATGATGATGATGATGATGATGATGATGAT 300  
 776 ATCATGGGCGAGGAGCGCTGTGAGACATTTTGTGACCTTGTGCTGCTGCTGCTGCTGCTGCT 835  
 301 ATCATCTCCAGACGTCAGACTTCCAGACCTTGTGGAGAGAACCTGGCGGGCGAG----- 354  
 836 AAGTACTAGAGGACAAGCAAGATCATCTTGGCTGATGAACGAGCGCGACGACCTC 895  
 355 -----TACAAAGATACGACCTGCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 408  
 896 GACATTGAGATCTGGGCGGACAGCTGCCAAAGGTCGTCAGTCCGAAAGCGCGC 955  
 409 GACCAGGATCTGCTGCTGAACCTCAACAGCAGCATTAAGGCGATCCGCGCGCGAGCT 468  
 956 GCCACCTCGACATGATCTCTCTGCGCCGACCAACTTTGCGAGCGTGCAGACGATGTG 1015  
 469 GCGACCGACGAGTACATCTTCTGCGAAGGCACTCTGAGACCGCGCTGAGCGTGGTGC 528  
 1016 TCCACTGGGAGCGGAGGCGCTCGGAGATTAACGACCGGATGGAACACCGCATTTG 1075  
 529 -----GACGTCAACGACACATAGAGATTTTACCGACCCCGAAGACAG 573  
 1076 CTGTACTTGTATGTCACAGATATCTGCACATCAACAACTCCGGTTCGACGCGGAGTGC 1135  
 574 ATGCTATGAAATGACAGCTGACGATCGGAGCTTCGCGGCACTTCGAGAGACTGC 633  
 1136 ACCAGACACAGCTC-----GACGCTTCAACGACTTCCGCGGACGCTGCTGAGGACAGAC 1189  
 634 GTGTCCGAGACATCGGAAAAGGCGGCTCAGGAGCTACACAGTGTGCTGAAGGACAAAT 693  
 1190 AAGCGCAGGACCATCTCTCCGAAAGCGGCGCTCCATGGAACCTTCGTCATGACTGCC 1249  
 694 AAGAGGTGGCTTACAGGCAATATGCCGGGGTTCCAAATGATGATGCTGGAAGTGC 753  
 1250 TTCTGCGCCAGAACAGAGGCAATTAAGCAAAACAGCGAGCTTACATTTGGCTTTGTGGC 1309  
 754 GTGTGCGGGATGCTGAGTACATGGCGAATTAACACGAGCTATGGAAGGTGGCTGTGG 813  
 1310 TGGGGTCCGGGCGAGCTTTGACAGCTGCTGATCTTGACTTGACTCCC 1358  
 814 TGGGAGCGGGCCATGTGGGAGAGACTACTATTTTCAGCATGTGAGGCC 862

RESULT 13  
 AAD20928  
 ID AAD20928 standard; DNA; 1008 BP.  
 AC AAD20928;  
 XX  
 DT 15-JAN-2002 (first entry)  
 DE Talaromyces emersonii beta-glucanase CEA DNA.  
 XX  
 KW Beta-glucanase; CEA; antileptic; fungicide; hyperlipemia; animal feed;  
 KW fabric; edible foodstuff; textile; brewing; distilling; biomethanation;  
 KW dental hygiene; leather-treatment; paper manufacture; plant extract;  
 KW detergent treatment; baking; ds.  
 XX  
 OS Talaromyces emersonii.

XX Key Location/Qualifiers  
 FH 1.1008  
 FT CDS  
 FT /tag- a  
 FT /product- "T. emersonii CEA protein"

MO200170998-A1.  
 27-SEP-2001.  
 20-MAR-2001; 2001MO-EP03174.  
 20-MAR-2000; 2000BP-0302263.  
 (STAM ) DSM NV.  
 Van Den Hombergh JPTW, Van Der Laan J, Daran JG, Herweijer MA,  
 Teufel DP;  
 WPI: 2001-648392/74.  
 P-PSDB; AAE12786.

New polypeptide for treating hyperlipemia and/or high serum  
 cholesterol and triglyceride levels, comprises the beta-glucanase  
 protein obtainable from Talaromyces fungus

Claim 5; Page 64-65; 76pp; English.

The invention relates to a beta-glucanase polypeptide obtainable from a  
 fungus of the genus Talaromyces, e.g. T. emersonii, having endoglucanase  
 activity (EC 3.2.1.4). Beta-glucanase polypeptide is useful for the  
 manufacture of a medicament for treating hyperlipemia, high serum  
 cholesterol and triglyceride levels. Beta-glucanase polypeptide are  
 useful for treating fungal or plant materials (plant pulp, plant  
 extracts), edible foodstuffs or ingredients, or fabrics, textiles or  
 clothes containing plant materials. Beta-glucanase polypeptide is  
 useful for reducing the viscosity of a plant material, for cleaving  
 beta-D-glucan polymers in the plant material, and for processing plant  
 pulp, juice or extract, by incubating the pulp, juice or extract.  
 Beta-glucanase polypeptide is applicable in brewing, distilling,  
 biomethanation, dental hygiene, leather-treatment, paper manufacture,  
 textile treatment or manufacture, baking or bread making, washing or  
 detergent treatment, treating flower bulbs or in animal feed.  
 CC beta-glucanase polypeptide is also useful during the production of milk  
 CC substituted from soybean. The present sequence is Talaromyces emersonii  
 CC beta-glucanase CEA DNA.  
 XX  
 SQ Sequence 1008 BP; 220 A; 298 C; 304 G; 186 T; 0 other;

Query Match 4.3%; Score 79; DB 22; Length 1008;  
 Best Local Similarity 47.8%; Pred. No. 1.8e-09;  
 Matches 355; Conservative 0; Mismatches 355; Indels 33; Gaps 3;

584 CATTTGCGGAGACGAGCGCTCAAGCTTTTGCATATCCGCTACATGACGATTGTC 643  
 199 CAGATCTCTATCGACAGGCGATGAACATCTTCCGCTGCTGATGAGCGCATG 258  
 644 CTCAACACACGCTGAGCGGCAAGCTGAGACCTCAACTGGGGCTCTTACACAGGTC 703  
 259 GTGCCCAACGATGAGCGGGCGGATGATCGGCTATTTCCAGGCGCTACACCGAGTT 318  
 704 GTCAAGCGCTGTCTGAGACGCGGCTACTGATGATGATGATGATGATGATGATGATGATGATGAT 763  
 319 ATCACTACTATTAACGACGCTGCGGTGCGCATATGACCCGATTAATCTTGGGCGCA 378  
 764 TACAAGCGGCGATCATCGGCCAGGAGCGGTGTCGAGACGACATCTTGTGACGCTGCG 823  
 379 TACTACAACAATATCATCTCC-----TCCCGCTGACCTTCCAGACTTCTCG 426  
 824 GTCAAGATGCAAAAGTACACGAGCAACGACAAATATCTTTGGCTGATGAACGAG 883  
 427 CACACTATGCGTCCAACTTTGGCGATTAATGACAATGTATTTTCACAGCAACAGCA 486



OY 897 A---CATGAGATCTGGGCGACAGCTGCCAAAAGGTCTCACTCGCATCCGAAAGCCG 953  
 DB 676 ATTCATATCCCGAGTGGGGAGACAGTCTCAATAGCTCGACAGCCGCTTGCGGGCCG 735  
 OY 954 GCGCCACCTCCGAGATGATCTCTCTCCCGGAAACCACTTTGCCAGCTCGACAGCTANG 1013  
 DB 736 GCTCGA---CGAACTACCTCTCTTACCGGTTCTTCTGGGGTACCGCAGCAGCTCC 792  
 OY 1014 TGTCCATGCGACGCGGGAAGCCCTGGGCAAGTATGCAACCGCGATGGAGACCGATT 1073  
 DB 793 CGACCCAGGCGGACCTTA---CCTCTCCAAATCACTGATCTTCTGGGGTACCAACA 849  
 OY 1074 TCGTACTCTTGTATGTCACAAAGTATCTGCATCAACAACCTCCGGGTCGACGCCGAGT 1133  
 DB 850 AACTCATTTTCATGTGACCAAGTACCTCGACAGCAACAACAGTGGACACCACTCCAACT 909  
 OY 1134 GCACCAACACAAAGCTGACAGCTTCAAGACCTTCCGGAGCTGGCTGAGGACAAA---CA 1190  
 DB 910 GCGTTACGAACAACACTGGTGTCTTTCACAGCAGCTGACCTGGCTCCACAGCAATGGCA 969  
 OY 1191 AGCGCCAGGCGCATCTCTCGAAACGGGCGCTGCATGGAACCTTGTCATGACTGCCCT 1250  
 DB 970 ACCGTAGGCGGCTTCTGACGAGACCTGTGAGAGTACTCTGACAGAGTGGGAGACAT 1029  
 OY 1251 TCTGCGCCGAGAAAGCCATTAGCGAAAGAGCAGCTCTACATT---GCCTTTGTGG 1307  
 DB 1030 ATGTTGCCCAAGAGCTTGATTCGTTCAAGCCAAAGATATACATTTGGCGGCTTGCCA 1089  
 OY 1308 GCTGGGGTCCCGGACACTTTGACACAGCTGTAATCTTTGACTCTGACTCCCTTG 1361  
 DB 1090 TCTGGGCGCCAGGTGATTCGACAGCAATACGCTCTTACCGCTACCCCGACAG 1143

## RESULT 10

AAFI4886

ID AAF14886 standard; cDNA: 923 BP.

AC AAF14886;

DT 13-MAR-2001 (first entry)

DE Trichoderma reesei EST SEQ ID NO:7409.

Multiple gene expression; filamentous fungal cell; EST;  
 expressed sequence tag; Fusarium venenatum; Aspergillus niger;  
 Aspergillus oryzae; Trichoderma reesei; identification; recombination;  
 culture condition; environmental stress; spore morphogenesis;  
 metabolic pathway engineering; catabolic pathway engineering; ss.

OS Trichoderma reesei.

PN WO200056762-A2.

PD 28-SEP-2000.

PF 22-MAR-2000; 2000WO-US07781.

PR 22-MAR-1999; 99US-0273623.

PA (NOVO) NOVO NORDISK BIOTECH INC.

PA (NOVO) NOVO NORDISK AS.

PI Berta RM, Rey MM, Shuster JR, Kauppinen S, Clausen IG, Olsen PB;

DR WPI: 2000-594572/56.

Monitoring differential expression of genes in filamentous fungal cells  
 uses fluorescence-labeled nucleic acids isolated from the cells and a  
 substrate of expressed sequence tags -

PS Claim 89; Page 2991-2992; 3161pp; English.

XX The present invention describes a method for monitoring differential

expression of genes in a first filamentous fungal (FF) cell relative to  
 expression of the same genes in one or more second filamentous fungal  
 cells. The method uses fluorescence-labeled nucleic acids isolated from  
 the FF cells and a substrate of expressed sequence tags (EST). The ESTs  
 are used in the methods for monitoring differential expression of genes  
 in a first filamentous fungal (FF) cell relative to expression of the  
 same genes in one or more second filamentous fungal cells. Monitoring  
 the global expression of genes from FF cells allows the production  
 potential of the microorganisms to be improved. New genes may be  
 discovered, possible functions of unknown open reading frames can be  
 identified and gene copy number variation and stability can be  
 monitored. The expression of genes can be used to study how FF cells  
 adapt to changes in culture conditions, environmental stress, spore  
 morphogenesis, recombination, metabolic or catabolic pathway  
 engineering. Using ESTs provides several advantages over genomic or  
 random cDNA clones including elimination of redundancy as one spot on an  
 array equals one gene or open reading frame, and organization of the  
 microarrays based on function of the gene products to facilitate  
 analysis of the results. AAF1248 to AAF1247 represents ESTs from  
 CC Fusarium venenatum; AAF11854 to AAF11853 represents ESTs from  
 CC niger; AAF14879 to AAF14878 represents ESTs from Aspergillus oryzae; and  
 CC AAF14879 to AAF15337 represents ESTs from Trichoderma reesei, which are  
 all specifically claimed in the present invention.

SQ Sequence 923 BP; 205 A; 255 C; 226 G; 228 T; 9 other;

Query Match

6.4%; Score 116.2; DB 21; Length 923;

Best Local Similarity 55.0%; Pred. No. 1,1e-18; Matches 270; Conservative 0; Mismatches 206; Indels 15; Gaps 2;

OY 454 GATCAATATCTGGGCTGCCATTCGCCGAATCGATTGGCTGCGACATGACGGCAG 513  
 DB 430 GGTCCGATTTGGCGGCTTAACATCGGGGTTTGACTTGGCTGACACAGATGGCAC 489  
 OY 514 CTGTCCGACTGACACGCTGCTGTGCCCTCTGAGCTACAAAGAGAGA----- 562  
 DB 490 TTGCGTTACCTCGAAGATTATCTCCGTTGAAGAACTTCACCGCTCAAAACATACCC 549  
 OY 563 -GATGGGCGCGCCAGATGAAAGCATTTGCGGAGAGCAGCGGCTCAAGCTTTTGCAAT 621  
 DB 550 CGATGGCATCGCCAGATGACGACATTCGTCAACGACGACGGATGACTATTTTCCGCTT 609  
 OY 622 ATCCGCTACATGAGCAGTTGCTCTCAACACAGCGGTGAGCGGCAAGCTGACAGCTCAA 681  
 DB 610 ACTGTGCGATGAGTGCAGTACTCGTCAACACAAATTTGGGGGGAATCTGATTCACAGAG 669  
 OY 682 CTGGGGCTCTTCAACAAAGGTGCTCAAGCCTGTCTGAGACGGGCGCTACTGATGAT 741  
 DB 670 CATTTCCAAAGTATGATCAGCTTCTTTCAGGGGCTCTGCTTNTGCGGCATACGATCGT 729  
 OY 742 TGACATGACACATTTGCGCGGCTACAAAGCGGCGCATCATCGGCGAGGAGCGGTGCGA 801  
 DB 730 CGACATCCACAAATTAAGCTGATGAGACGGGGGATCATTTGTTCANMGCGCCCTACTAA 789  
 OY 802 CGACATCTTTGTGACCTCTGGGTCCAGATGCGAAAGTACTAGAGAGA---CAACGACAA 858  
 DB 790 TGTCTCAATTCACAGAGCTTTGTGCGCAAGTTGGCATCAAAAGTACGATTTAATGANG 849  
 OY 859 GATCATCTTTGGCGCTGATGACGAGCGCAGCACTGCGACATTTGAGATCTGGCGCAGAC 918  
 DB 850 GGNNGGTTGCGATNATGAATGAACCCGACGAGTGAACATTAACACCTGGGGTTGCNC 909  
 OY 919 GTGCCAAAGG 929  
 DB 910 GGGCTAANMG 920

## RESULT 11

AAFI4904

ID AAF14904 standard; cDNA: 584 BP.

AC AAF14904;

XX

CC (AAW02034). The CBH1, CBH11, EGI and EGI1 enzymes have catalytic  
 CC core domains useful for reducing dye redeposition (backstaining) on  
 CC cellulose-contg. fabrics such as denim, whilst maintaining or  
 CC increasing abrasion during stonewashing. Truncated enzymes comprising  
 CC these catalytic core domains can be obtd. by inserting the appropriate  
 CC DNA fragment into a vector and using this to transform a trichoderma  
 CC sp. host cell.

XX Sequence 1849 BP; 444 A; 507 C; 440 G; 458 T; 0 other;

Query Match 15.1%; Score 275.6; DB 17; Length 1849;

Best Local Similarity 59.2%; Pred. No. 6.1e-58; Matches 489; Conservative 0; Mismatches 334; Indels 3; Gaps 1;

QY 563 GATGCGCCCGCCAGATGATGATTCGCGCAAGACAGCGGCTCAACGTCCTTCGATA 622  
 DB 832 GATGCGATCCGCGCGATGCGCACTTCGTCACAGAGAGCGGATGACTATTTCCGCTTA 891  
 QY 623 TCCGCTACATGGCAGTTTGTCTCAACACAGGTTGAGCGCAAGCTGAGAGCTCAGC 682  
 DB 892 CTTGTCGATGCGAGTACCTGCTCAACACAAATTTGGCGCATTTGATTCACAGAGC 951  
 QY 683 TGGGGCTCTACACAGGTCGTCAGGCTGTCTCAGAGGCGGCTCTGATGATTT 742  
 DB 952 ATTTCAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1011  
 QY 743 GACATGACACACTTTCGCGCTACACAGCGGCGATCATGCGCGAGGAGCGGTGCGAC 802  
 DB 1012 GACATGACACACTTTCGCGCTACACAGCGGCGATCATGCGCGAGGAGCGGTGCGAC 1071  
 QY 803 GACATGCTTTCGCGCTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGT 862  
 DB 1072 GCTCAATTCAGAGGCTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1131  
 QY 863 ATCTTGGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 922  
 DB 1132 TGGTGGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1191  
 QY 923 CAAAAGGTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 982  
 DB 1192 CAAAAGGTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1251  
 QY 983 GGAACCAACTTTCGCGCTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCG 1042  
 DB 1252 GGAATGATGCGATGCTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCG 1311  
 QY 1043 AAGATTGACACCGGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1102  
 DB 1312 CAAATGACGAAACCGGATGATGATGATGATGATGATGATGATGATGATGATGATG 1371  
 QY 1103 GACATCAACACTTCGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCG 1159  
 DB 1372 GACATCAACACTTCGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCG 1431  
 QY 1160 AAGACTTTCGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCG 1219  
 DB 1432 TCTCCGCTTTCGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGT 1491  
 QY 1220 GCGTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1279  
 DB 1492 GGTGGCAACCTTCGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1551  
 QY 1280 AACAGGAGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1339  
 DB 1552 AACTCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1611  
 QY 1340 ATCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1385  
 DB 1612 GTCTGACGAAACCGGATGATGATGATGATGATGATGATGATGATGATGATGATG 1657

RESULT 9  
 AAA53338

ID AAA53338 standard; cDNA; 1372 BP.  
 XX  
 AC AAA53338:  
 XX  
 DT 25-SEP-2000 (first entry)  
 XX  
 DE Endoglucanase encoding nucleotide sequence.  
 XX  
 KM Endoglucanase; cellulose hydrolysis; ss.  
 XX  
 OS Corticium rolfsii.  
 XX  
 FH Key Location/Qualifiers  
 FT CDS 27..1199 /tag- a  
 FT /product- "Endoglucanase"  
 FT sig\_peptide 27..77 /tag- b  
 FT /note- "Putative signal sequence"  
 FT mat\_peptide 78..1196 /tag- c  
 FT /note- "Putative mature protein"  
 XX  
 PN JP2000106887-A.  
 PD 18-APR-2000.  
 XX  
 PF 30-SEP-1998; 98JP-0377864.  
 XX  
 PR 30-SEP-1998; 98JP-0377864.  
 XX  
 PA (YASO/) YASOKAWA D.  
 XX  
 DR WPI: 2000-402850/35.  
 DR P-PSDB: AAB03663.  
 XX  
 PT A gene encoding endoglucanase  
 PS  
 XX  
 Example; Page 7; 10pp; Japanese.  
 CC The present sequence represents an endoglucanase encoding gene. The  
 CC invention relates to the gene and the protein encoded by it, having  
 CC endoglucanase activity. Also included in the invention is an expression  
 CC vector containing the gene sequence, and a fungus such as Aspergillus  
 CC transformed by the expression vector. Endoglucanase is used to hydrolyse  
 CC cellulose.  
 XX  
 SQ Sequence 1372 BP; 331 A; 388 C; 305 G; 348 T; 0 other;

Query Match 8.2%; Score 150; DB 21; Length 1372;  
 Best Local Similarity 54.1%; Pred. No. 5.8e-27;  
 Matches 419; Conservative 0; Mismatches 340; Indels 15; Gaps 5;

QY 597 ACGAGGCTTCAACGCTTTCGCTATTCGCTACATGCGAGCTTGTCTCAACACAGG 656  
 DB 376 ACGAGGCTTCAACGCTTTCGCTATTCGCTTTCGCTTTCGCTTTCGCTTTCGCTTTCG 435  
 QY 657 TGCAGGCTTCAACGCTTTCGCTATTCGCTTTCGCTTTCGCTTTCGCTTTCGCTTTCG 716  
 DB 436 GTGATGATTTACCAACCTTTCGCTTTCGCTTTCGCTTTCGCTTTCGCTTTCGCTTTCG 495  
 QY 717 TCGAGAGGCGGCTTTCGCTTTCGCTTTCGCTTTCGCTTTCGCTTTCGCTTTCGCTTTCG 776  
 DB 496 TGCACACGCGGTTCGCTTTCGCTTTCGCTTTCGCTTTCGCTTTCGCTTTCGCTTTCG 555  
 QY 777 TCAATGCGCGAGGAGGCGGTGCGAGCAATCTTTCGCTTTCGCTTTCGCTTTCGCTTTCG 836  
 DB 556 TCAATGCGCGAGGAGGCGGTGCGAGCAATCTTTCGCTTTCGCTTTCGCTTTCGCTTTCG 615  
 QY 837 AGTACTAGAGCAACGAGATCATCTTTCGCTTTCGCTTTCGCTTTCGCTTTCGCTTTCGCT 896  
 DB 616 CTTACTAGAGCAACGAGATCATCTTTCGCTTTCGCTTTCGCTTTCGCTTTCGCTTTCGCT 675

```

FT CDS 765..1689.
FT /tag= f
FT /product= catalytic core domain
FT /note= "seq id no 15"
XX
XX W09516782-A.
XX
XX 22-JUN-1995.
XX
XX 19-DEC-1994; 94WO-US14163.
XX
XX 17-DEC-1993; 93US-0169948.
XX
XX (GENEW ) GENENCOR INT INC.
XX
XX Clarkson KA, Collier KD, Fowler T, Larenas E, Ward M;
XX WPI. 1995-231574/30.
XX P-PSDB; AAR77264.
XX
XX Pure, truncated fungal cellulase protein from Trichoderma - useful to
XX reduce or eliminate dye, colourant or pigment back-staining or
XX redeposition in stone-washing or bio-polishing
XX
XX Claims 35, 58, 59; Figure 4; 105pp; English.
XX
XX Figure 4 depicts the genomic DNA and AA sequence of EgII derived
XX from T. longibrachiatum. DNA fragments are claimed which comprise
XX SEQ ID nos 11, 19 and 15; 15 and 19; and 15, 19 and 11.
XX Genes for EgI and EgII have been isolated from T. longibrachiatum
XX and the protein domain structure has been confirmed (Penttila, M.
XX et al., 1986, Gene, 45, 253-263; Van Arsdell, J.N. et al., 1987,
XX Bio/Technology, 5, 60-64; and Saloheimo, M. et al., 1988, Gene
XX 63, 11-21).
XX
XX Sequence 1849 BP: 444 A; 507 C; 440 G; 458 T; 0 other;
XX
XX Query Match 15.1%; Score 275.6; DB 16; Length 1849;
XX Best Local Similarity 59.2%; Pred. No. 6.1e-58;
XX Matches 489; Conservative 0; Mismatches 334; Indels 3; Gaps 1;
XX
XX 563 GATGGCGCGCCGATGAGCATTTTGGCCGGAAGACGAGCGGCTTCAAGCTTTTGGCATA 622
XX ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
XX DB 832 GATGGCATGGCGAGATGAGCATTTTGGCCGGAAGACGAGCGGCTTCAAGCTTTTGGCATA 891
XX ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
XX QY 623 TCCGCTACATGGAGCTTGTCTCCACACACAGCGGAGCGCAAGCTGAGCGAGCTCAAC 682
XX ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
XX DB 892 CCGTGTGGATGGCAGTACCTCGTCACACACAAATTTGGCGGCAATCTTGATTCACAGAGC 951
XX ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
XX QY 683 TGGGGCTCTACACAAAGGTCTGCAAGCGCTGTCTGAGAGCGGCGCTTACTGATGATT 742
XX ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
XX DB 952 ATTTCCAAAGTATGATGAGCTTGTTCAGGGGTGCTGTCTGTGGCGCATCTGATCGATC 1011
XX ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
XX QY 743 GACATGACAACTTTGCCCGCTACAAAGCGGCGCATCATGGCCAGCGAGCGTGTGGAC 802
XX ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
XX DB 1012 GACATTCACAAATTTGCTGATGAGAAAGCGGTGGGATCATTTGTCAGGCGCCCTACTAAT 1071
XX ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
XX QY 803 GACATCTTTGTGCAAGCTTGGGTGCGATTCGCAAAAGTACTACAGGAGCAACAGCAAGATC 862
XX ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
XX DB 1072 GCTCAATTCAGAGCTTTTGGTGTGCGAGTTGGCATCAAAAGTACCATCTCAGTCAGAGGTG 1131
XX ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
XX QY 863 ATCTTTGGCTGATGAAAGCGCGGCGAGCATCTGAGATGAGATCTGGCGGAGAGGTGC 922
XX ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
XX DB 1132 TGGTGGGATCATGATAGAGCCCGACGAGCGTGAACATCAACACCTGGGTGCGACAGGTG 1191
XX ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
XX QY 923 CAAAAGGTGCTACGTGCGATCCGAAAGCGCGGCGCATCTGCGAGATGATCTCTGCC 982
XX ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
XX DB 1192 CAAGAGGTGTAAACCGCAATCGCAACGCTGGGTGCTAGCTGCGAATTCATCTTTGCGCT 1251
XX ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
XX QY 983 GGAACCACTTTGCCAGCTCGAGACGTATGTCTCCACTGGCAGCGCGGAGAGCCCTCGGC 1042
XX ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
XX DB 1252 GGAATATGATTGGCAATCTCTGGGGCTTTCATATCGAGTGGCAGTGCAGCGCCCTGTCT 1311

```

```

QY 1043 AGATTACGAAACCCGATGAGAACACCGATTTGCTGTACTTTGATGTCACAAAGTATCTC 1102
XX ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1312 CAGTGCAGAACCCGATGGGTGTCACACAGCAATCTGATTTTTCAGGTGCAAAATCTTG 1371
XX ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1103 GACATCAACAACTTCGGGTGCGACGCCGAGTGCACACAGCAACAC--GTCGACGCTTC 1159
XX ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1372 GACTCAGACAACTCCGCTACTCAGCCGAAATGACTACAAATTAATTTGACGGCGCTTT 1431
XX ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1160 AACGACTTGGGAGTGGGTGAGGAGCAACACGAGCGGCGATCTCTCGAAGAGGGC 1219
XX ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1432 TCTCCGCTTGCACCTTGGCTCCGACAGAACATTCGAGGCTATCTGACAGAACCGGT 1491
XX ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1220 GCGTCCATGGAACCTTCGTGCATGACTGCTTTCGCGCCGAGAACAGGCATTAGCGAA 1279
XX ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1492 GGTGGCAAGCTTCATGATGCTGCAATCAAGACATGTGCGACGAATATCTCAACCGAG 1551
XX ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1280 AACAGGACGCTTACATTTGGCTTTGGCGTGGGGTCCGCGACGCTTTGACACGCTGTAC 1339
XX ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1552 AACTCAGATGTCTATCTTGTGCTATGTGGTGGGTGGCGGATCTTTGATGACAGGTAT 1611
XX ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1340 ATCTTGACTGTGACTCCCTCGGCAAGCCCGGCACTACACGAGCA 1385
XX ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1612 GTCTTGACGGAACCCGACTACAGTGTACTCATGACAGGACA 1657
XX ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
RESULT 8
AAT32223
ID AAT32223 standard; DNA: 1849 BP.
XX
XX AAT32223;
AC
AC AAT32223;
XX
XX 28-OCT-1996 (first entry)
DE
DE Trichoderma endoglucanase II gene.
XX
XX Trichoderma endoglucanase II gene.
XX
XX Endoglucanase II; EgII; cellulase; cellulose; denim;
XX stonewashing; dye redeposition; backstaining; ss.
XX
XX Trichoderma longibrachiatum.
XX
XX Location/Qualifiers
XX FT exon 1..590
XX FT /tag= a
XX FT /codon_start= 262..264
XX FT Intron 591..764
XX FT /tag= b
XX FT exon 765..1849
XX FT /tag= c
XX
XX W09623928-A1.
XX
XX 08-AUG-1996.
XX
XX 29-JAN-1996; 96WO-US00977.
XX
XX 01-FEB-1995; 95US-0382452.
XX
XX (GENEW ) GENENCOR INT INC.
XX
XX Clarkson KA, Collier KD, Fowler T, Larenas E, Ward M;
XX WPI. 1996-371466/37.
XX P-PSDB; AAM02032.
XX
XX Treatment of cellulose-contg. fabrics such as denim, e.g.
XX stone-washing - using truncated cellulase enzyme to increase
XX abrasion and give reduced redeposition of dye
XX
XX Disclosure; Fig 4A-4C; 124pp; English.
XX
XX Genomic DNA sequences (AAT32220-24) of Trichoderma longibrachiatum
XX respectively code for cellobiohydrolase I (CBHI) (AAM02022), CBHI
XX (AAM02025), endoglucanase I (EgI) (AAM02029), EgII (AAM02032) and EgIII

```

```

FT mat_peptide 535..590
FT /tag- b
FT /note- "endoglucanase-II coding region of exon 1"
FT mat_peptide 766..1689
FT /tag- c
FT /note- "endoglucanase-II coding region of exon 2"
FT misc_feature 433..534
FT /tag- d
FT /note- "linker"
FT misc_feature 325..432
FT /tag- e
FT /note- "cellulose binding domain"
XX W09516360-A1.
XX 22-JUN-1995.
XX 19-DEC-1994; 94NO-EP04212.
XX 17-DEC-1993; 93US-0169948.
XX (FINN-) FINNFEEDS INT LTD.
XX (GEMV) GENECOR INT INC.
XX Bedford MR, Clarkson KA, Collier KD, Fowler T, Larenas E;
XX Morgan AJ, Ward M;
XX MPI; 1995-231296/30.
XX P-PSDB; AAR79540.
XX New animal feed additive - comprising one or more endo:glucanase
XX enzymes and opt. a cello:bio:hydrolase enzyme.
XX Disclosure; Fig 2A-C; 92pp; English.
XX Endoglucanase-II (EG-II) is used in an animal feed-additive to
XX produce a cereal-based animal feed, especially for fowl. EG-II
XX may be produced recombinantly, optionally without the
XX cellulose-binding domain. The feed-additive can be incorporated
XX into a cereal-based feed (barley, wheat, triticale, rye and maize)
XX where it improves the conversion ratio and/or increases the
XX digestibility of the feed. The feed-additive enables a
XX conventional cereal-based feed to be modified by reducing its
XX energy, protein and/or amino acid content while simultaneously
XX maintaining the same nutritional levels of energy, protein and
XX amino acids available to the animal. The feed-additive also
XX contains cellobiohydrolase, and optionally a xylanase, protease,
XX mannanase or a phytase.
XX Sequence 1849 BP; 444 A; 507 C; 440 G; 458 T; 0 other;
SQ
Query Match 15.1%; Score 275.6; DB 16; Length 1849;
Best Local Similarity 59.2%; Pred. No. 6,1e-58;
Matches 489; Conservative 0; Mismatches 334; Indels 3; Gaps 1;
XX 563 GATGGCCGCGGAGATGAAGATTTGCGGAGACGAGCGGCTCAAGCTTTGCGATA 622
XX ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
XX 832 GATGGCATTCGCGAGATGACACTTGTCAACGAGACGGGATGACTATTTCCGCTTA 891
XX ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
XX 623 TCGGCTACATGGAGCTTGTCCCAACACGSGTGAGCGGACGCTGAGAGCTCAAC 682
XX ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
XX 892 CCTGTGCGAGTGGAGTACCTGTCAACACATTTGGCGCAATCTTGATTCAGAGC 951
XX ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
XX 683 TGGGGCTCTAACAACAAGTGTCAACGCTGTCTGAGAGCGGCGCTACTGATGATT 742
XX ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
XX 952 ATTTCAGATATGATGCTGTTTCAGGGGTGCTGTCTGGCGCATCTGATGCTGTC 1011
XX ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
XX 743 GACATGCAACACTTGGCCCTCAACGCGGCGCATCATCGCCAGGAGCGGTGCGAC 802
XX ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
XX 1012 GACATCCACATTAATGCTCATGAGGATGGGATCATTTGTTCAGGCGCCCTAAT 1071
XX ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
XX 803 GACATCTTTGACCTCTGTGGTCCAGATGCAAGATACGAGACGACAAAGATC 862
XX ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

```

Db 1072 GCTCAATTCAGAGCCCTTTGGTCCAGTGGCATCAAGTACGATCTCAAGGGTG 1131
Qy 863 ATCTTGGCCCTGATGAAGAGCCCGACAGCTCGACATTTGATGTGGCGACAGCTGC 922
Db 1132 TGCTTGGCATCATGATGATGAGCCCGACAGCTGATCAACACCTGAGCTGCGAC 1191
Qy 923 CAAAGGCTGTCACTGCGGATCGGAAGCGCGCCGACCTGCGAGATGATCTCTGCCC 982
Db 1192 CAAGAGGTGTAAACCGCAATCCGACAGCTGTGCTAGTGTGCAATTCATCTTTGGCT 1251
Qy 983 GGAACCACTTGGCCAGCTGTGAGACGATATGTTCACATGCGGAGGAGGAGCCCTGGC 1042
Db 1252 GGAATGATTTGGCAATCTGTCTGGGCTTTCATATTCGATGCGACGCGCTGTCT 1311
Qy 1043 AAGATTAAGACCCGATGAAGACCGCATTTGCTGTACTTGAATGTCACAGATATCTC 1102
Db 1312 CAAGTCAGAACCCCGATGGGTCAACAGCATGTGATTTTGGACGTCACAAATACTTG 1371
Qy 1103 GACATCAACAACCTCGGCTGCGACCGCGAGTGACACACAGACAC--GTGAGCGCTTC 1159
Db 1372 GACTCAGACACTCCGGTACTACGCGCAATGTACTACAAATTAACATTTGACGCGCTTT 1431
Qy 1160 AAGCACTTGGGAGCTGGCTGAGGACAGACAGCGCCGACGATCATCTCCGAAGCGGC 1219
Db 1432 TCTCCGCTTGGCACTTGGCTCGACAGAACATGCGCAGGCTATCTGACAGAAACGGT 1491
Qy 1220 GCGTTCATGAACCTTGTGATGATGACTGCTTGTGCGCCGAGAACAGGCGCATTAGCGAA 1279
Db 1492 GGTGGCAACGTTCACTGCTGTGATACAGACATGTGCGCAACATTCATATCTCAACGAG 1551
Qy 1280 AACAGCAGCTGTACATTTGCTTGTGGGCTGGGGTGGCGGACGCTTGTGACAGCTGCTAC 1339
Db 1552 AACTCAGATGCTATCTTGTGATTTGTTGGTGGGCGGATCATTTGATAGACGAT 1611
Qy 1340 ATCTTGAATCTGATCCCTCGGCAAGCGCGGCAACATCACCGACA 1385
Db 1612 GTCTGACGGAACACCGATGACAGTGAATGATGAGGAGACA 1657
XX
XX RESULT 7
XX ID AA091284 standard; DNA: 1849 BP.
XX AC AA091284;
XX AC 13-DEC-1995 (first entry)
XX DE T. longibrachiatum endoglucanase EgII.
XX XX Cellulase; cellulose; signal; catalytic core; cellulase binding;
XX KW linker; ss.
XX OS Trichoderma longibrachiatum.
XX FH Key Location/Qualifiers
XX FT CDS 262..324
XX /tag- a
XX /product- signal
XX /note- "seq id no 31"
XX FT CDS 325..432
XX /tag- b
XX /product- cellulose binding domain
XX /note- "seq id no 7"
XX FT CDS 433..534
XX /tag- c
XX /product- linker
XX /note- "seq id no 23"
XX FT CDS 535..590
XX /tag- d
XX /product- catalytic core domain
XX /note- "seq id no 15"
XX FT Intron 591..764
XX /tag- e

```

QY 1280 AACAGCGAGCTACATGGCTTTGGGGTGGGGTCCGGCAGCTTTGACACGCTGTAC 1339  
 DB 1470 AACTGAGATGTATCTATCTGCTATGCTGGGTGGGGTCCGCTCATTTTATACACTTAT 1529  
 QY 1340 ATCTTACCTGACCTCCCTCGGCAAGCCGGGAACTACACGCAACAAGCTCATGAAC 1399  
 DB 1530 ATCTCTACGGAAGAACGCTACTGGAAGCGGTAACTCGTGGACGACACATCCCTAGTAGC 1589  
 QY 1400 GAGTG 1404  
 DB 1590 TCGTG 1594

# RESULT 5 ABL60701

ID ABL60701 standard; DNA: 1720 BP.

AC ABL60701;

DT 27-AUG-2002 (first entry)

DE T. viride CMC nucleotide sequence.

KW Cellulase; endoglucanase; surfactant; detergent; cellulose; paper;  
 KM pulp treatment; CMC; ds.

OS Trichoderma viride.

PN WO200238754-A1.

PD 16-MAY-2002.

PF 12-NOV-2001; 2001WO-JP09858.

PR 10-NOV-2000; 2000JP-0343921.

PA (MEIJ) SEIKA KAISHA LTD.

PI Koga J, Nakane A, Baba Y, Kono T;

DR WPI; 2002-47155/50.

PT Cellulase preparations containing transconjugant-originated  
 endoglucanase and non-ionic surfactants, useful in detergent  
 compositions, in treating cellulose fibers and delinking waste paper and  
 improving freeness of paper pulp -

PS Disclosure: Page 35-36; 38pp; Japanese.

CC The invention relates to a cellulase preparation comprising a  
 CC transconjugant-originated endoglucanase and a non-ionic surfactant. The  
 CC endoglucanase is selected from RCEI, RCEII, MCEI, MCEII or PCEI  
 CC proteins. The preparations are useful in detergent compositions, in  
 CC treating cellulose fibers and delinking waste paper and improving the  
 CC freeness of paper pulp. The fibers treated by the preparations have  
 CC reduced feathering and improved skin-feel and appearance with colour  
 CC clarification, local change in colour and softening, and after delinking  
 CC and paper pulp treatment, there is an improvement on freeness of the  
 CC paper pulp. This treatment with the cellulase preparation can be operated  
 CC at significantly lower cost. The present sequence represents the  
 CC T. viride CMC nucleotide sequence.

CC Sequence 1720 BP; 390 A; 470 C; 410 G; 450 T; 0 other;

Query Match 15.9%; Score 289.8; DB 24; Length 1720;

Best Local Similarity 59.8%; Pred. No. 1.8e-61;

Matches 505; Conservative 0; Mismatches 337; Indels 3; Gaps 1;

QY 563 GATGGCGCGGAGATGATTCGGCGAAGCAGCGGCTCAACGCTTTGGCAT 622  
 DB 750 GACGGATATGGCGGAGATGACGACCTTCGTCAACGATGATGGGATGCTATTTCCGCTA 809  
 QY 623 TCCGCTACATGGCAGTTTCTCTCAACAACACGCGTGGAGGCAAGCTGACGAGCTCAAC 682

DB 810 CCCGTGGATGGCAGATACCTGTAACAACAACTGGGTGGAACCTGCTCAGTCCACAGT 869  
 QY 683 TGGGGCTCTCACAACAAGGTCGTAACGCGTGTCTGGAAGAGGGGCCCTCTCATGATT 742  
 DB 870 ATCTCGAATATATAGCTGCTGTTCAAGGGGTGCTCTCTCTCTATATACATCATC 929  
 QY 743 GACATGCACAACTTTGGCCGCTACAAAGCGGCGATCATCGGCGAGGAGCGTGTGGAC 802  
 DB 930 GACATGCACAACTTATGCTGCTGATGGAACGGTGGAAATCATTTGGCCAGGAGCCCTACAAAT 989  
 QY 803 GACATCTTTGTGACCTTGGGTGCGAATCGCAAGTACTAGGAGCAACAAGCAAGATC 862  
 DB 990 GCCAGATTTCACAGTCTTTTGGTGGCGAGTTGGCATTCGAAGACCGCTCAGTGGAGGTG 1049  
 QY 863 ATCTTTGGCTGATGAACGAGCGGCAAGCAAGCTTGACATTTGAGTCTGGGGGAGAGTGC 922  
 DB 1050 TGGTTGGGAATTAATGATGAGGCCCAAGAGTGAACATCAACTTGGGCTGACGCGTT 1109  
 QY 923 CAAGAAGTGTGCTACTGCGATCCGAAGCGCGGCGCACTCGGAGATCTCTGCCC 982  
 DB 1110 CAAGAGTGTGCTACTGCGATCCGAAGCGCGGCTGCTACGTCGGAATATTTCTGTGCC 1169  
 QY 983 GGAACCACTTTGCGCAGCGTGCAGAGCTATGTGTGCACTGGCAGCGCGGAAGCCTCGGC 1042  
 DB 1170 GGAATGATTTATCAATCTGGCGCAGCTTTTATTTCCGATGGCAGTCCAGCGCCCTGTCT 1229  
 QY 1043 AAGATTACGAACCGGATGGAACGACGATTTCTGATCTGATGTCGCAAGTATCTC 1102  
 DB 1230 CAGGTACGAACCCCTGATGGATCAACAGCAATCTTAATCTTGATGCTCCAGAACTTCA 1289  
 QY 1103 GACATCAACAACCTCGGGGTGCGACGCGGAGTGCAGCAGCAAGCAAGCTGCA---CGCCTTC 1159  
 DB 1290 GACTCGGAGCAACTCCGGTACTCAGCGGATGATGACTACAAACAACATCGAGGCGCTTT 1349  
 QY 1160 AACGACTTGGCGGAGCTGCTGAGGCAAGCAAGCGCGAGGCTCATCTCCGAACGGGC 1219  
 DB 1350 GCTCTCTGCGCCACTTGGCTTGCAGCAAGCAACCGCGAGCTTATCTGAGGGAACCGGC 1409  
 QY 1220 GCGTCCATGGAACCTTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 1279  
 DB 1410 GGTGCAATGTTAGTCTGCTGCTCAAGATTTTGGCAACGATGCTCAACGACG 1469  
 QY 1280 AACGAGGAGCTACATTTGGCTTTGGGCTGGGTCGGGCACTTTGACACGCTGTAC 1339  
 DB 1470 AACTGAGATGTATCTTGGCTATGCTGGGTGGGTCGGGCTCATTTGATACACTTAT 1529  
 QY 1340 ATCTTACCTGACCTCCCTCGGCAAGCCGGGCACTACACGCAACAAGCTCATGAAC 1399  
 DB 1530 ATCTCTACGGAAGAACGCTACTGGAAGCGGTAACTCGTGGACGACACATCCCTAGTAGC 1589  
 QY 1400 GAGTG 1404  
 DB 1590 TCGTG 1594

# RESULT 6 AA097716

ID AA097716 standard; DNA: 1849 BP.

AC AA097716;

DT 14-JAN-1996 (first entry)

DE Endoglucanase-II genomic DNA sequence.

KW Endoglucanase-II; cellulase complex; feed-additive; ss.

OS Trichoderma longibrachiatum.

FT key Location/Qualifiers

FT sig\_peptide 262..324  
 FT /\*tag= a



CC The present sequence encodes a cellulase enzyme SCE-3. The cellulase  
 CC is used in the methods of the invention for treatment of  
 CC cellulose-containing fibres, for bleaching denim-dyed  
 CC cellulose-containing fibres, for eliminating fluffs from  
 CC cellulose-containing fibres, for weight loss treatment of  
 CC cellulose-containing fibres and of desecelylated triacetate  
 CC rayon, all by contacting the preparation with such fibres.  
 CC It is useful in the textile and related industries.  
 XX  
 XX

SO Sequence 1720 BP; 390 A; 470 C; 410 G; 450 T; 0 other;

Query Match 15.9%; Score 289.8; DB 20; Length 1720;  
 Best Local Similarity 59.8%; Pred. No. 1.8e-61;

Matches 505; Conservative 0; Mismatches 337; Indels 3; Gaps 1;

563 GATGGGGCGGGCGAGATGATTTGCGGAGAGAGAGCGGCTCAACGCTTTCCGATA 622  
 DB 750 GAGCGATGCGCGAGATGCGAGCTTCGTCACAGATGATGGATGATTTTCCGCTA 809  
 623 TCCGCTACATGCGAGTTTGTCTCAACACACGCTGAGCGGCAAGCTGACAGCTCAAC 682  
 DB 810 CCGGTGGAGTGGAGATACCTCGTAACACATCTGGTGGAACTCTCGATTCCACCACT 869  
 683 TGGGGCTCTACAAACAGGTGTCAACGGCTGTCTGAGACGGGCGCTACTGATGATT 742  
 DB 870 ATCTCGAAGTATGATGAGCTCTTCAAGGGGCTGCTCTCCGTGATGATGATCATC 929  
 743 GACATGACACATTTTCCGCTACAAAGCGGCGGATATGCGGAGAGCGGCTGTGAGAC 802  
 DB 930 GACATCCACATATATCTCGATGAGCGGTGATGATGATGCGGAGAGCGGCTTACAAAT 989  
 803 GACATCTTTGTGACCTGTGGTGTGAGATGCAAGATGATGATGATGATGATGATGATG 862  
 DB 990 GCCCATTTTACAGT 1049  
 863 ATCTTTGGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 922  
 DB 1050 TGGTGTGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1109  
 923 CAAGAAGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 982  
 DB 1110 CAAGAGGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1169  
 983 GGAACCACTTTTCCAGCGTGTGATGATGATGATGATGATGATGATGATGATGATGATG 1042  
 DB 1170 GGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1229  
 1043 AGATTTACGAAACCGGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1102  
 DB 1230 CAGGTAAAGCAACCTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1289  
 1103 GACATGACACATGCGGGTGTGAGCGGCGGATGATGATGATGATGATGATGATGATGAT 1159  
 DB 1290 GACTCGGACATCTCGGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1349  
 1160 AACGACTTGTGCGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1219  
 DB 1350 GCTGCTGTGCGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1409  
 1220 GCGTCCATGGAACCTTGTGATGATGATGATGATGATGATGATGATGATGATGATGATG 1279  
 DB 1410 GGTGGCAAGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1469  
 1280 AACAGGAGGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1339  
 DB 1470 AACTGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1529  
 1340 ATCTTACTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1399  
 DB 1530 ATTCTGAGGAAAGGCTTACTGATGATGATGATGATGATGATGATGATGATGATGATG 1589  
 1400 GAGTG 1404  
 |||

DB 1590 TCGTG 1594

RESULT 3

ABL59234

ABL59234 standard; DNA; 1720 BP.

ABL59234; (first entry)

07-OCT-2002

Nucleotide sequence of a Trichoderma viride gene sequence.

Detergent; cellulase; endoglucanase; cellulose binding domain; CBD;

glycol copolymer; dirt removal; colour clarification; feather removal;

stone washed; ss.

Trichoderma viride.

Key Location/Qualifiers

Inttron 500..682

JP2002142760-A.

21-MAY-2002.

10-NOV-2000; 2000JP-0343925.

10-NOV-2000; 2000JP-0343925.

(MEIJ) MEIJI SEIKA KAISHA LTD.

WPI; 2002-561125/60.

Detergent compositions, useful for treatment of textiles, e.g. colour

clarification, comprise cellulase preparations with endoglucanase and

terephthalic acid-alkylene glycol copolymer or terephthalic

acid-oligoalkylene glycol copolymer

Disclosure; Page 8-9; 10pp; Japanese.

The specification describes detergent compositions, which are composed

of cellulase preparations comprising an endoglucanase with a cellulose

binding domain (CBD) and one or more terephthalic acid-alkylene glycol

copolymer or terephthalic acid-oligoalkylene glycol copolymer. The glycol

copolymer is of a formula given in the specification. The detergent

compositions provide smooth removal of dirt. The detergent compositions

of the invention are used for treatment of textiles for colour

clarification, removal of feathers or improvement of texture to give a

stone washed appearance. The present sequence represents a Trichoderma

viride gene sequence, which is used in the course of the invention.

Sequence 1720 BP; 390 A; 470 C; 410 G; 450 T; 0 other;

Query Match 15.9%; Score 289.8; DB 24; Length 1720;

Best Local Similarity 59.8%; Pred. No. 1.8e-61;

Matches 505; Conservative 0; Mismatches 337; Indels 3; Gaps 1;

563 GATGGCGCGGGCGAGATGATTTGCGGAGAGAGAGCGGCTCAACGCTTTCCGATA 622  
 DB 750 GAGCGATGCGCGAGATGCGAGCTTCGTCACAGATGATGGATGATTTTCCGCTA 809  
 623 TCCGCTACATGCGAGTTTGTCTCAACACACGCTGAGCGGCAAGCTGAGAGCTCAAC 682  
 DB 810 CCGGTGGAGTGGAGATACCTCGTAACACATCTGGTGGAACTCTCGATTCCACCACT 869  
 683 TGGGGCTCTACAAACAGGTGTCTCAACGGCTGTCTGAGAGCGGCGCTACTGATGATT 742  
 DB 870 ATCTCGAAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 929  
 743 GACATGACACATTTTCCGCTACAAAGCGGCGGATGATGATGATGATGATGATGATGATGAT 802  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

XX 13-SEP-1996; 96JP-0243695.  
PA (MEIJ ) MEIJI SEIKA KAISHA LTD.  
PI Aoyagi K, Moriya T, Murakami T, Sumida N, Watanabe M;  
XX WPJ: 1998-250959/22.  
DR P-PSDBJ; AAM57421.  
XX  
PT Regulatory sequence for Trichoderma viride derived cellulase cbhl  
CC gene - for producing Humicola insolens derived endo-glucanase  
CC  
CC Disclosure: Pages 52-55; 92pp; Japanese.  
PS  
XX  
CC The present sequence appears in the specification. The specification  
CC describes a new regulatory sequence for Trichoderma viride derived  
CC cellulase cbhl gene and the establishment of a system for mass producing  
CC cellulase in molds such as T. viride. As the regulatory objective of  
CC cbhl genes originating in T. viride can highly express objective  
CC proteins, proteins such as cellulase can be expressed. An expression  
CC vector containing the regulatory sequence and Humicola insolens derived  
CC endo-glucanase NCE4 DNA was produced, and used to produce endo-glucanase  
CC at 15 grams per litre.  
XX  
SQ Sequence 1463 BP; 334 A; 401 C; 367 G; 361 T; 0 other:  
Query Match 15.9%; Score 289.8; DB 19; Length 1463;  
Best Local Similarity 59.8%; Pred. No. 1.7e-61;  
Matches 505; Conservative 0; Mismatches 337; Indels 3; Gaps 1

OY	563	GATGGCGGCGGCAGATACGATTTCGCCCAAGCAGCGGCTTCAACTCTTTGGCATA	622
Db	593	GACGGTATCGGCCCAATCATCACACTTGTCGAACGATGATGGGATGACTATTTCCGCCCTA	652
OY	623	TCCGCTACATGAGCATTTGTCTCTTAACAACACGGTGGAGCGGAGGTGAGAGGCTCAAC	682
Db	653	CCCCTCGGATGGCACTACTCGTAAACAACATCTGGGGAGAACHTCTGATTCACCACT	712
OY	683	TGGGCTCTTACACACAGGCTGTCAAAGCCTGTCTTCGAGACGGGCGCTACTGCATGATT	742
Db	713	ATCTCGAAGTATGATACACTCGTTTAGGGGTGCTCTGTCTCTGCTATATACGATCATC	772
OY	743	GACATGCACTATTGGCCGCTACACAGCGGCGCATCTGGCCAGGAGAGCGGTGCGAC	802
Db	773	GACATCCCACTATTAGCTTCGATGAGAACGGTGAATCATATTGGCCAGGAGGCCCTTA	832
OY	803	GACATCTTTGTGCAGCTCTGGGCTCCAGATCGCAAGTACTACGAGACACAACAGAATC	862
Db	833	GCCAGTTACCAGTCTTTGGTCCGAGTTGGCATCGAATGACGCGCTCACTGAGAGCG	892
OY	863	ATCTTTGGCCGATGAAGACGCCGACGACCTCGACATTTGAGATCTGGGCGCAGAGCTGC	922
Db	893	TGCTTCGGAATTAATGAATGAGCCGCCACGACATCAACACTTTGGCTCCACGGATT	952
OY	923	CAAAAGTGCTCACTGCGATCCGAAAAGCGCGGCGCACCTCGACAGATGATCTCTCGCC	982
Db	953	CAGAAGGTCGTCACTGCAATCGCAAGCCGGGTCTAGCTGCCAATATTTCTCTGCT	1012
OY	983	GGAACCAACTTGGCAGCGTGGAGACGATGTGTCCACTGCGAGCGGAGAGCCCTCGGC	1042
Db	1013	GGAATGATTAATCAATCTGCGGACCTTTATTTCCTGAGTGGCAGTGCAGCCCGCTGTCT	1072
OY	1043	AAGATTACGACCCGGATGAGAGACCCGATTTGCTGTACTTTGATGTCCACAAGTATC	1102
Db	1073	CAGTAAACGAACTCGATGATGATCAACAACGAATCTAATCTTGATGTCCACAAGTACTTA	1132
OY	1103	GACATCAACAACCTCCGGGTGCACGCGAGTGACACACAGCAACAGCTGA---CGCCTTC	1159
Db	1133	GACTCGACAACCTCCGGTACTCAGCGCAATGACTACTAACAACATCATGACGAGCGCTTT	1192
OY	1160	AAGCACTTGGGAGCTGGCTGAGGAGAAACAAGCGCCAGGCGATATCTCCGAACGGGC	1219
Db	1193	GCTCTCTCGCACTTGCTTGCACAGAACACCGCGGCTATTCTCAGGAAACCGGC	1252

Oy	1220	GCGTCATGGAACCTTCCGTGCATGACTGCCCTTCTGCGCCAGAACAAAGCCATTAGCGAA	1279
Db	1253	GGTGGCATACTTGTAAGTCTCTGCATCCCAAGATTGTGCCCACAAGATCCAGTACTCAACCAG	1312
Oy	1280	AACAGCAGCTGTCAATTGGCTTTGTGGGCTGGGGGTGGCGGCAAGCTTTGACACGTGCTAC	1339
Db	1313	AACTCAGATGCTCATCTTCTGCTAAGTCAGTCGGGCTGGCGGGTCCGGTTCATTGTATAGCACCTTAT	1372
Oy	1340	ATCTTGACTCTGACTCCCTCCGTGGGAGCGCCGCAACTRACACCGCAACAGCTCATGAC	1399
Db	1373	ATTCTGACGAAAGCCCTACTGTGAAGCGGTAACTCTGTGGAGGACACATCCCTTAGTTAGC	1432
Oy	1400	GAGTG 1404	
Db	1433	TCTGTG 1437	
 RESULT 2 AAK15437 standard; DNA: 1720 BP.			
XX	AAK15437:		
AC	AAK15437:		
XX	07-MAY-1999 (first entry)		
DE	DNA encoding a cellulase enzyme designated SCE-3.		
XX			
KW	Cellulase enzyme; cellulose-containing fibre; bleaching; denim-dyed;		
KM	fluff elimination; weight loss treatment; deacetylated triacetate rayon.		
KW	SCE-3; ds.		
OS	Trichoderma viride.		
XX			
FH	key	Location/Qualifiers	
FH	CDS	171..1610	
FT	/tag= a		
FT	/product= cellulase		
FT	/note= "contains 1 intron"		
FT	exon	171..499	
FT	/tag= b		
FT	sig_peptide	171..233	
FT	/tag= c		
FT	mat_peptide	234..1610	
FT	/tag= d		
FT	intron	500..682	
FT	/tag= e		
FT	exon	683..1610	
FT	/tag= f		
 WC0854332-A1.			
PX	03-DEC-1998.		
XX			
PD	27-MAY-1998; 98WO-JP02326.		
XX			
PF	27-MAY-1997; 97JP-0137258.		
XX			
PR	(MEIJ ) MEIJI SEIKA KAISHA LTD.		
PA			
XX	Aoyagi K., Koga J., Kono T., Murakami T., Nakamura Y;		
PI	Sato Y., Sumida N., Watanabe M;		
XX			
DR	WP1: 1999-070218/06.		
XX	P-PSDB; AAW97208.		
XX			
PT	Cellulose preparation containing highly active cellulase SCE3 -		
PT	e.g. in treating cellulose-containing fibres to enable fluff		
PT	elimination, weight loss and bleaching, and in weight loss treatment		
PT	of deacetylated tri:acetate rayon		
XX			
XX	Disclosure; Date 27-30; 44pp; Japanese.		
XX			

GenCore version 5.1.4-p5.4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 15, 2003, 14:03:24 ; Search time 379.376 Seconds  
(without alignments)  
10839.251 Million cell updates/sec

Title: US-10-028-245-1  
Perfect score: 1826  
Sequence: 1 gtcgaccacgcytcgcgttc.....aatcaaaaaaaaaaaaaa 1826

Scoring table: IDENTITY\_NUC  
Gapop 10.0, Gapext 1.0

Searched: 2185239 seqs, 112599159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 08  
Maximum Match 1008

Listing first 45 summaries

Database :

N.Geneseq.101002.\*  
1: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT:\*  
2: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT:\*  
3: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT:\*  
4: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT:\*  
5: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT:\*  
6: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT:\*  
7: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT:\*  
8: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT:\*  
9: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT:\*  
10: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT:\*  
11: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT:\*  
12: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT:\*  
13: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT:\*  
14: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT:\*  
15: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT:\*  
16: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT:\*  
17: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT:\*  
18: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:\*  
19: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT:\*  
20: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:\*  
21: /SID2/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:\*  
22: /SID2/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:\*  
23: /SID2/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT:\*  
24: /SID2/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	289.8	15.9	1463	19	AAV29597
2	289.8	15.9	1720	20	AAV15437
3	289.8	15.9	1720	24	ABL59234
4	289.8	15.9	1720	24	ABL43252
5	289.8	15.9	1720	24	ABL60701
6	275.6	15.1	1849	16	AAQ97716
7	275.6	15.1	1849	16	AAQ91284
8	275.6	15.1	1849	17	AAT32223
9	150	8.2	1372	21	AAA53338

10	116.2	6.4	923	21	AAF14886	Trichoderma reesei
11	101.4	5.6	584	21	AAF14904	Trichoderma reesei
12	79.4	4.3	912	19	AAV16436	Glucanase II gene.
13	79	4.3	1008	22	AAV20928	Talaromyces emersoni
14	74.2	4.1	985	24	ABQ44816	Oligonucleotide fo
15	74.2	4.1	985	24	ABQ44817	Oligonucleotide fo
16	55.6	3.0	390	13	AAQ21833	Randomising oligon
17	55.6	3.0	390	14	AAQ36859	PCR primer for 5'
18	55.6	3.0	390	22	AAV76910	Sequence containin
19	55.6	3.0	390	24	AAV72775	Oligo #7 for clon
20	53.8	2.9	1241	17	AAV13948	Bacillus VIP2A(a)
21	53.8	2.9	1241	17	AAV13956	Bacillus VIP2A(a)
22	53.8	2.9	1241	18	AAV74007	Maize optimised-VI
23	53.8	2.9	1241	18	AAV74008	Maize optimised-B.
24	53.8	2.9	1241	19	AAV16184	Maize optimised-B.
25	53.8	2.9	1241	19	AAV16186	Maize optimised-DN
26	53.8	2.9	1358	17	AAV13949	Bacillus VIP2A(a)
27	53.8	2.9	1358	18	AAV74009	Maize optimised-B.
28	53.8	2.9	1358	19	AAV16188	Maize optimised-DN
29	53.8	2.9	1389	17	AAV13946	Maize optimised-VI
30	53.8	2.9	1389	18	AAV74002	Maize optimised-Ba
31	53.8	2.9	1389	19	AAV16175	Maize optimised-DN
32	53.8	2.9	1399	17	AAV13953	Maize optimised-VI
33	53.8	2.9	1399	18	AAV74000	Maize optimised-B.
34	53.8	2.9	1399	19	AAV16173	Maize optimised-DN
35	53.8	2.9	4031	17	AAV13947	Maize optimised-VI
36	53.8	2.9	4031	18	AAV74010	Maize optimised-B.
37	53.8	2.9	4031	19	AAV16189	Maize optimised-DN
38	52.2	2.9	1198	18	AAV63043	Aspergillus niger
39	52.2	2.9	1198	18	AAV60427	Cellulase cDNA clo
40	50.6	2.8	5163	19	AAV20700	Cryptosporidium pa
41	50.6	2.8	5163	21	AAV61849	ORF encoding a por
42	50.6	2.8	5163	24	ABT04778	C parvum GP900 gen
43	50.6	2.8	5318	19	AAV20701	Cryptosporidium pa
44	50.6	2.8	5318	21	AAA61848	DNA encoding a por
45	50.6	2.8	5318	24	ABT04777	C parvum GP900 gen

#### ALIGNMENTS

RESULT 1  
ID AAV29597 standard; DNA; 1463 BP.  
AC AAV29597;  
XX  
DT 02-SEP-1998 (first entry)  
XX  
DE DNA sequence of the specification.  
XX  
KW Regulatory sequence; cellulase chl gene; mass production;  
KW Humicola insolens; endo-glucanase NCE4; ds.  
XX  
OS Trichoderma viride.  
XX  
FH Key Location/Qualifiers  
FT CDS 14..1453  
FT /tag= a  
FT /note= "contains an intron"  
FT sig\_peptide 14..76  
FT /tag= b  
FT mat\_peptide 77..1450  
FT /tag= c  
FT /tag= d  
FT Intron  
PN W09811239-A1.  
PD 19-MAR-1998.  
XX  
PF 16-SEP-1997; 97WO-JP03268.  
XX

NTMDQTLVLIDNQAIDGIRACATSOYIFAEKGSWSGAMTWADINDMKALTLPQDK  
LVEFMHOYLDSDSGSTGCVSEFTIGAEKLOAAQWLKDNKXVILLEGYAGAGADVC  
TAISMELEYANNTDVMKGAVMWTAGPMADYMFMSPEPSPAISGMLDVEPTLG"

```

/db_xref="GI:9107955"
/translation="WVSTPRRIEESMLDASDERDAGCFGWAVOUDOPSRALLEIA
IALSRMTRRGVAVDGLTGDCGILLIRDPALFLRAAGELACIMGSPRAVAVMEVPE
EBAANAHCQVLEAELOQVYVLCGRREPPIRNSCGOARLOTLPHEVEVDIDPKH
SDADAVFLFKRRAROLCDYHDYVATLSPVALSTGMYLPDLPLFYFDLOSE
LSSSVIVHQRSTNTLPRMPLAHPRLANGELINTIEGNRMAQARSKVQSPED
MAEPDVISMKSDSODNMLLELJAGMDLLOALIRLIPATRTLEFEDLASFY
EEYGLNTEPMWDGPAIVADGYAACMDRNGRLRPARMLTDHRLHVESEGAVMLP
VERVIRKGLGEGEMATDLKRGDILSDAIDRINARA.PYKOMLOOQVTLTOLIED
PBLVEEPSDOLCSYOKLFOJLSTEBEVVLPPLAETEBEATSGMKDDPMAYLSOI
RPLDTFROAEQVNTNPIDPLRESSTAMSLTETGNETIEHAGETVHVIDLSEVL
SQKRLQKMPQVLEKNRDLIDLSYVEEGLOAKRMCDEDEOARDOIWILLSDR
YVPERPAVHALIATGAVHNLKRVGLRDLVNL1ETGTAHPHNAVCGLTGATVAV
PYLSOTLENRGRGLKLOKAGEOSQIRSKYKLSK1ISMGGCTLSYGA
OLFEIVGLDPPVVALCEPPTPARIGVSIMOLHTECELTARANDRIRAPYEGILKY
VHGGERHYNPPVMTLOHATRTGDAAHRRMDVHTTPPBARLLOLKAETITP
LDHVAAPDVLCRFDTAATSLGALSPHEALAVANRIGGNSGEGEDPARVTE
RRSKIKOVASGRFVPEYLINAENVQIRIAGAKREGGOLGKRVNDLIRLRYKA
PGIGLISPPHHDIYSIEDLAQILDQVNPQALSVLVSHVGTIAAGVAKGA
DLITVSGHDGSGASPISSIRYAGVWEVAVHOALHELEBANDYVKKGAGGLVVR
PRGAFPEARHTPIIGNTCLYGATGELFACAGVGEFLEANGALVIEGAGDHCEY
MDGVVLVYKGVNFGAGFTGGLAYIIDIEDEVDYVNHEDILHIVSEGFENRO
HLRHLVVRRAALGSIWAGQIIDEFPRDIYGRKFPVKPAASIESLASLSRSA"
8277. 8576
/gene="Xf2711"
8277. 8576
/gene="Xf2711"
/translation="hypothetical protein; identified by sequence
similarity: putative; ORF located using Glimmer/RBSfinder"
/codon_start=1
/translation_table=11
/product="hypothetical protein"
/protein_id="AAF85508.1"
/db_xref="GI:9107956"
/translation="MLETMAFYDTGKATHYRLLPMNALKNPHIRLNSCSSTAT
HRYTCLTMRGSSNMTMDVHLFEYTELHMHMLMPTTHHIAITPHRPSMQO"
8690. 9607
/gene="Xf2712"
8690. 9607
/gene="Xf2712"
/translation="similar to SP/P22008 (percent identity: 49 %/query
alignment coverage: 86.9 %/subject alignment coverage:
97.1 %); identified by sequence similarity: putative; ORF
located using Glimmer/RBSfinder/Start codon shift: -123"
/codon_start=1
/translation_table=11
/product="pyrroline-5-carboxylate reductase"
/protein_id="AAF85509.1"
/db_xref="GI:9107957"
/translation="MOHRMTRATATMLASFSSIAMPHVSLPSPPIAFIGGNAR
SMIAGMIRGAPSTISICVETPIAALDALIADENIKAVTATEAATDSTVAVAPQ
VMDACITPLPVPQPTPLVISAIGITCKLHMGDTCAIRAMPNPAIIGACV
GEPATSHTAOTRLLAEELFENSVGKMTLTERALMDAVTAAGSGPAPVFLLEAMEA
AIIQOISTANARPIVQCTLLGAARMTLIGETPALNRHRTVSPHGTAAIATFQEN
DPQALVAKAIDATRRARALGKANDTEPT"
BASE COUNT 2749 a 3234 c 2131 g 1883 t
ORIGIN

Query Match 4.4%; Score 80.4; DB 1; Length 9997;
Best Local Similarity 48.5%; Pred. No. 2e-05; Indels 9; Gaps 2;
Matches 287; Conservative 0; Mismatches 296;

QY 732 ACTGATGATTGACATCACTTGGCCGCTAACAAAGCGGCATCATCGCCAGAG 791
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 499 ACATCATTTAGATGTCACAACTACTCCGAATATAAGATGATGCGACATGACA 558
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 792 GCGTGTGGACGACATCTTTGCGACCTCTGCTCCAGATCGCAAGACTACGAGACA 851
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 559 ATGTCCCTATCAGCGCCTTGCAGATCATGAAAGCTGTGCTCCATACGTTTGCACAG 618
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

/db_xref="GI:9107955"
/translation="WVSTPRRIEESMLDASDERDAGCFGWAVOUDOPSRALLEIA
IALSRMTRRGVAVDGLTGDCGILLIRDPALFLRAAGELACIMGSPRAVAVMEVPE
EBAANAHCQVLEAELOQVYVLCGRREPPIRNSCGOARLOTLPHEVEVDIDPKH
SDADAVFLFKRRAROLCDYHDYVATLSPVALSTGMYLPDLPLFYFDLOSE
LSSSVIVHQRSTNTLPRMPLAHPRLANGELINTIEGNRMAQARSKVQSPED
MAEPDVISMKSDSODNMLLELJAGMDLLOALIRLIPATRTLEFEDLASFY
EEYGLNTEPMWDGPAIVADGYAACMDRNGRLRPARMLTDHRLHVESEGAVMLP
VERVIRKGLGEGEMATDLKRGDILSDAIDRINARA.PYKOMLOOQVTLTOLIED
PBLVEEPSDOLCSYOKLFOJLSTEBEVVLPPLAETEBEATSGMKDDPMAYLSOI
RPLDTFROAEQVNTNPIDPLRESSTAMSLTETGNETIEHAGETVHVIDLSEVL
SQKRLQKMPQVLEKNRDLIDLSYVEEGLOAKRMCDEDEOARDOIWILLSDR
YVPERPAVHALIATGAVHNLKRVGLRDLVNL1ETGTAHPHNAVCGLTGATVAV
PYLSOTLENRGRGLKLOKAGEOSQIRSKYKLSK1ISMGGCTLSYGA
OLFEIVGLDPPVVALCEPPTPARIGVSIMOLHTECELTARANDRIRAPYEGILKY
VHGGERHYNPPVMTLOHATRTGDAAHRRMDVHTTPPBARLLOLKAETITP
LDHVAAPDVLCRFDTAATSLGALSPHEALAVANRIGGNSGEGEDPARVTE
RRSKIKOVASGRFVPEYLINAENVQIRIAGAKREGGOLGKRVNDLIRLRYKA
PGIGLISPPHHDIYSIEDLAQILDQVNPQALSVLVSHVGTIAAGVAKGA
DLITVSGHDGSGASPISSIRYAGVWEVAVHOALHELEBANDYVKKGAGGLVVR
PRGAFPEARHTPIIGNTCLYGATGELFACAGVGEFLEANGALVIEGAGDHCEY
MDGVVLVYKGVNFGAGFTGGLAYIIDIEDEVDYVNHEDILHIVSEGFENRO
HLRHLVVRRAALGSIWAGQIIDEFPRDIYGRKFPVKPAASIESLASLSRSA"
8277. 8576
/gene="Xf2711"
8277. 8576
/gene="Xf2711"
/translation="hypothetical protein; identified by sequence
similarity: putative; ORF located using Glimmer/RBSfinder"
/codon_start=1
/translation_table=11
/product="hypothetical protein"
/protein_id="AAF85508.1"
/db_xref="GI:9107956"
/translation="MLETMAFYDTGKATHYRLLPMNALKNPHIRLNSCSSTAT
HRYTCLTMRGSSNMTMDVHLFEYTELHMHMLMPTTHHIAITPHRPSMQO"
8690. 9607
/gene="Xf2712"
8690. 9607
/gene="Xf2712"
/translation="similar to SP/P22008 (percent identity: 49 %/query
alignment coverage: 86.9 %/subject alignment coverage:
97.1 %); identified by sequence similarity: putative; ORF
located using Glimmer/RBSfinder/Start codon shift: -123"
/codon_start=1
/translation_table=11
/product="pyrroline-5-carboxylate reductase"
/protein_id="AAF85509.1"
/db_xref="GI:9107957"
/translation="MOHRMTRATATMLASFSSIAMPHVSLPSPPIAFIGGNAR
SMIAGMIRGAPSTISICVETPIAALDALIADENIKAVTATEAATDSTVAVAPQ
VMDACITPLPVPQPTPLVISAIGITCKLHMGDTCAIRAMPNPAIIGACV
GEPATSHTAOTRLLAEELFENSVGKMTLTERALMDAVTAAGSGPAPVFLLEAMEA
AIIQOISTANARPIVQCTLLGAARMTLIGETPALNRHRTVSPHGTAAIATFQEN
DPQALVAKAIDATRRARALGKANDTEPT"
BASE COUNT 2749 a 3234 c 2131 g 1883 t
ORIGIN

RESULT 15
LOCUS AF054512 1240 bp mRNA linear PLN 31-MAR-1998
DEFINITION Aspergillus aculeatus endoglucanase V (cell) mRNA, complete cds.
ACCESSION AF054512
VERSION AF054512.1 GI:2997730
KEYWORDS
SOURCE
ORGANISM Aspergillus aculeatus.
Aspergillus aculeatus
Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.
REFERENCE 1 (bases 1 to 1240)
AUTHORS Kofod, L.V., Dalboe, H., Andersen, L.N., Kauppinen, S. and
Christgau, S.
TITLE Aspergillus aculeatus endoglucanase
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 1240)
AUTHORS Muller, S., Sandal, T., Kamp-Hansen, P. and Dalboe, H.
TITLE Direct Submision
JOURNAL Submitted (19-MAR-1998) Microbial Discovery I, Novo Nordisk A/S,
Novo Alle, Build. 1B1.20, Bagsvaerd 2880, Denmark
FEATURES
source
1. 1240
/organism="Aspergillus aculeatus"
/strain="WO 94/14953"
/db_xref="taxon:5053"
1. 1240
/gene="cell"
27. 1025
/gene="cell"
/EC_number="3.2.1.4"
/Note="Cell: belongs to glycosyl hydrolase family 5"
/codon_start=1
/product="endoglucanase V"
/protein_id="AAC08587.1"
/db_xref="GI:2997731"
/translation="MKLLNLVLAAGAAVAAPHEHTKRASVFEWISNNSDAEFG
TAIPGTWGDYIFPDPSAATLVSKGMNIFRYOFMERLVVPSMGSYDDAYLNNLT
VNAIAAGVHAIVDPHNGRYNNETISSTADPQFPMWIAQOFKNDLIVIFDTNNE

```

Db 1052 CGAATGTCGGATGCCGCCATCCGCTGCAACCGGTGCGGCAACCTGATCTTGTGCGCG 1111  
 QY 984 GAACCAACTTGGCAGCGTCGAGACGATGTGTGCATGCGCAGCGCGGAGACCCCTCGGCA 1043  
 Db 1112 GAAATCTCTGTGGGAGCACAATAGCTGGATGGCGAGACTATGTGGCGCCACAGCGCG 1171  
 QY 1044 AGATTAGAAC---CCGGATGGAAGACCGATTTGTCTGTACTTTGATGTGCACAGATATC 1100  
 Db 1172 TTGTCAATGCTCGGTGTGAAGATCCGCTGCATCTATATGCTTGCAGTCCACAGATATC 1231  
 QY 1101 TCGACATCAACAATCCGGTCCGACCCGAGTCA-----CCACAGACAGCTGACG 1154  
 Db 1232 TGAACGATGTCTTTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT 1291  
 QY 1155 CATTCAACGACTTCCGCGATGCTGTGAGGACGAACAAGCCGACCATCATCTCCGAA 1214  
 Db 1292 CATTCGAGACTACACGACATGCTGCTCGACACATGCGACAGCCGCTTACTGGCCAGT 1351  
 QY 1215 CGGCGCGCTCCATGGAACCTTGTGATGATGATGCTTGTGCGCCGACAGACAGCCATTA 1274  
 Db 1352 TCGCGCTGCGGAGATGAGACGCTGCTGAGGCTCTCAGTCCGATGCTGATACGCTCG 1411  
 QY 1275 GCGAAACACGACGATCTACATGCTTGTGCGCTGCTGCGCTGCGCGCGACGT 1325  
 Db 1412 AACGGGCGGAGATGCTGATGATGCTGATGCTTATTTGCGCGGCGAGTACT 1462

RESULT 14  
AE004077

LOCUS Xylella fastidiosa 9a5c, section 223 of 229 of the complete genome.  
 DEFINITION Xylella fastidiosa 9a5c, section 223 of 229 of the complete genome.  
 ACCESSION AE004077 AE003849  
 VERSION AE004077.1 GI:9107952

KEYWORDS  
 SOURCE Xylella fastidiosa 9a5c.  
 ORGANISM Bacteria; Proteobacteria; gamma subdivision; Xanthomonas group;  
 Xylella.

REFERENCE  
AUTHORS

1 (bases 1 to 9997)  
 Simpson, A.J.G., Rehnach, F.C., Arruda, P., Abreu, F.A., Acencio, M.,  
 Alvarado, A., Alves, L.M., Araya, J.E., Baia, G.S., Baptista, C.S.,  
 Barros, M.H., Bonaccorsi, E.D., Bordin, S., Bove, J.M., Britones, M.R.,  
 Bueno, M.R., Camargo, A.A., Camargo, L.E., Carriaro, D.M., Carier, H.,  
 Colauto, N.B., Colombo, C., Costa, F.F., Costa, M.C., Costa-Neto, C.M.,  
 Coutinho, L.L., Cristofani, M., Dias-Neto, E., Docena, C., El-Dorri, H.,  
 Facchini, A.P., Ferreira, A.J., Ferreira, V.C., Ferro, J.A.,  
 Fraaga, J.S., France, S.C., Franco, M.C., Frohme, M., Furian, L.R.,  
 Garnier, M., Goldman, G.H., Goldman, M.H., Gomes, S.L., Gruber, A.,  
 Ho, P.L., Hohnsels, J.D., Junqueira, M.L., Kemper, E.L., Kitajima, J.P.  
 and Marino, C.L.  
 The genome sequence of the plant pathogen Xylella fastidiosa. The  
 Xylella fastidiosa Consortium of the Organization for Nucleotide  
 Sequencing and Analysis  
 Nature 406 (6792), 151-157 (2000)  
 20365717  
 10910347

JOURNAL  
MEDLINE  
PUBMED  
REFERENCES  
AUTHORS

2 (bases 1 to 9997)  
 Simpson, A.J.G., Rehnach, F.C., Arruda, P., Abreu, F.A., Acencio, M.,  
 Alvarado, A., Alves, L.M.C., Araya, J.E., Baia, G.S., Baptista, C.S.,  
 Barros, M.H., Bonaccorsi, E.D., Bordin, S., Bove, J.M., Britones, M.R.S.,  
 Bueno, M.R.P., Camargo, A.A., Camargo, L.E.A., Carriaro, D.M.,  
 Carier, H., Colauto, N.B., Colombo, C., Costa, F.F., Costa, M.C.R.,  
 Costa-Neto, C.M., Coutinho, L.L., Cristofani, M., Dias-Neto, E.,  
 Docena, C., El-Dorri, H., Facchini, A.P., Ferreira, A.J.S.,  
 Ferreira, V.C.A., Ferro, J.A., Fraaga, J.S., France, S.C., Franco, M.C.,  
 Frohme, M., Furian, L.R., Garnier, M., Goldman, G.H., Goldman, M.H.S.,  
 Gomes, S.L., Gruber, A., Ho, P.L., Hohnsels, J.D., Junqueira, M.L.,  
 Kemper, E.L., Kitajima, J.P., Krieger, J.E., Kuramae, E.E., Laigret, F.,  
 Lambais, C.R., Leite, L.C.C., Lemos, E.G.M., Lemos, M.V.F., Lopes, S.A.,  
 Lopes, C.R., Machado, J.A., Machado, M.A., Madalena, A.M.B.N.,  
 Madalena, H.M.F., Marino, C.L., Marques, M.V., Martins, E.A.L.,  
 Martins, E.M.F., Matsukuma, A.Y., Menck, C.F.M., Miracca, E.C.,  
 Miyaki, C.Y., Monteiro-Vitorello, C.B., Moon, D.H., Nagai, M.A.,

FEATURES  
source

Nascimento, A.L.T.O., Netto, L.E.S., Nhani Jr., A., Nobrega, F.G.,  
 Nunes, L.R., Oliveira, M.A., de Oliveira, M.C., de Oliveira, R.C.,  
 Palmieri, D.A., Paris, A., Peixoto, B.R., Pereira, G.A.G., Pereira  
 Jr., H.A., Pesquero, J.B., Quaglio, R.B., Roberto, P.G., Rodrigues, V.,  
 de M. Rosa, A.J., de Rosa Jr., V.E., de Sa, R.G., Santelli, R.V.,  
 Sawasaki, H.E., da Silva, A.C.R., da Silva, F.R., da Silva, A.M., Silva  
 Jr., W.A., da Silva, J.F., Silvestri, M.L.Z., Siqueira, W.J., de  
 Souza, A.A., de Souza, A.P., Terezini, M.F., Truffi, D., Tsai, S.M.,  
 Tshako, M.H., Vailarda, H., Van Sluys, M.A., Verjovski-Almeida, S.,  
 Vettore, A.L., Zago, M.A., Zatz, M., Zeldin, J. and Setubal, J.C.  
 Direct Submission  
 Submitted (02-JUN-2000) Organization for Nucleotide Sequencing and  
 Analysis, Bioinformatics Lab - IC/Unicamp, C.P. 6176, Campinas, SP  
 13083-970, Brazil

## location/Qualifiers

1..9997  
 /organism="Xylella fastidiosa 9a5c"  
 /db\_xref="taxon:160492"  
 /clone="9a5c"  
 144..1214  
 /gene="Xf2708"  
 144..1214  
 /gene="Xf2708"  
 /note="similar to SP|P17974 (percent identity: 47 %/query  
 alignment coverage: 86.8 %/subject alignment coverage:  
 74.2 %) ; identified by sequence similarity; putative; ORF  
 located using Glimmer/RBSfinder"  
 /codon\_start=1  
 /transl\_table=11  
 /product="endo-1,4-beta-glucanase"  
 /protein\_id="AA85505.1"  
 /db\_xref="GI:9107953"  
 /translation="MPIISFLCPLICPSCFLILTLATTTTPASPSHATNLTAYGINL  
 PGPEFSNPKGVLYKDYLPYSASDVAIVVKKIMYRLVPLMERQPTLNDLDTQ  
 LKLIKAVGAKOKXILIDVNVSEYDELGTNPVSADIRAKRLSIOFANDX  
 AVTFGLMNEPHEITSTMAQQAQALNALSITACANNVLPPTAASGASHLSYIG  
 IPNGEALNINDPAKHAEEVHOYLNENSTGTGECISTIGAEKLEAFNWLRTIHK  
 TGFLEGFATGNDCTNOALEGMLSYIEENDVNLGLTWGSGNPWFQPDYFENLPSND  
 GSDKPMOSILSESHARKITKIKH"  
 complement(2196..3680)  
 /gene="Xf2709"  
 complement(2196..3680)  
 /note="similar to SP|P09832 (percent identity: 55 %/query  
 alignment coverage: 97.4 %/subject alignment coverage:  
 101.9 %) ; identified by sequence similarity; putative; ORF  
 located using Glimmer/RBSfinder"  
 /codon\_start=1  
 /transl\_table=11  
 /product="glutamate synthase, beta subunit"  
 /protein\_id="AA85506.1"  
 /db\_xref="GI:9107954"  
 /translation="NSRKHAFOFLDIPROMPORIPEVETSGMRELXGKDEKAEAY  
 OARCLDCGNPTCSKCPVHNAIPOMLOLVOERIDBAALCATNPDLREYGRVCPQ  
 DRLCESSCTLEEFGAVTIGAVERYIVDTALQDGRDLSHVQPTGRNVAVAGFPAIG  
 ACADRLARAGVAVVDRYEDQIGLQFGIPSPKRLKAVSTRKYLEGVGEFFRIGV  
 EIKGDIGLEALDEHYDAVFGVAGVARTDGLGGOGLKAVLPALPLVONGRIVSGND  
 PYGRPLAGMDVOLDPDPVGTGSVNVGGGTGDCVRSARIGATVATCTCYRDEANM  
 PGSAREVANREGVRLFENRPLSTOSGDKAICVTVETKLABDANGRRNAVPI  
 OGTREAVLEADVYIAGFSFPALEPMILTAOSTIGSNGRIVALSIPDGDLTOTTHPK  
 LFAAGDGVKRAADVIVAVAGRDAESIVAMLKATVQVMDVAVI"  
 complement(3821..8290)  
 /gene="Xf2710"  
 complement(3821..8290)  
 /note="similar to SP|P09831 (percent identity: 53 %/query  
 alignment coverage: 99.7 %/subject alignment coverage:  
 97.9 %) ; identified by sequence similarity; putative; ORF  
 located using Glimmer/RBSfinder/start codon shift: -117"  
 /codon\_start=-1  
 /transl\_table=11  
 /product="glutamate synthase, alpha subunit"  
 /protein\_id="AA85507.1"





repeat\_region predicted by Homology"  
/evidence-not\_experimental  
202. .218  
/note="RS05849  
HYPOTHETICAL PARA BOX  
predicted by Homology"  
/evidence-not\_experimental  
complement(227. .243)  
/note="RS05850  
HYPOTHETICAL PARA BOX  
predicted by Homology"  
/evidence-not\_experimental  
complement(263. .279)  
/note="RS05851  
HYPOTHETICAL PARA BOX  
predicted by Homology"  
/evidence-not\_experimental  
complement(307. .323)  
/note="RS05852  
HYPOTHETICAL PARA BOX  
predicted by Homology"  
/evidence-not\_experimental  
complement(374. .390)  
/note="RS05853  
HYPOTHETICAL PARA BOX  
predicted by Homology"  
/evidence-not\_experimental  
complement(516. .532)  
/note="RS05854  
HYPOTHETICAL PARA BOX  
predicted by Homology"  
/evidence-not\_experimental  
complement(670. .686)  
/note="RS05855  
HYPOTHETICAL PARA BOX  
predicted by Homology"  
/evidence-not\_experimental  
complement(733. .749)  
/note="RS05856  
HYPOTHETICAL PARA BOX  
predicted by Homology"  
/evidence-not\_experimental  
complement(853. .869)  
/note="RS05857  
HYPOTHETICAL PARA BOX  
predicted by Homology"  
/evidence-not\_experimental  
complement(919. .935)  
/note="RS05858  
HYPOTHETICAL PARA BOX  
predicted by Homology"  
/evidence-not\_experimental  
1219. .2244  
/gene="repa"  
/note="Rsp0001; RS02238"  
1219. .2244  
/function="elements of external origin; plasmid-related  
functions non-bacterial functions"  
/note="Product confidence : probable  
Gene name confidence : probable  
predicted by Codon\_usage  
predicted by Homology  
predicted by Framed"  
/codon\_start=1  
/evidence-not\_experimental  
/transl\_table=1  
/product="PROBABLE REPA REPLICASE PROTEIN"  
/protein\_id="CAD17152.1"  
/db\_xref="GI:17430468"  
/translation="MGRITVPSGCGVIATPEKORLFDPEAQAEREDAMOSGEVGL  
SRASVOYTLPRAPKSPWTRSSGNISLIMIOGPTOORSEKASNGRORIVSETVS  
GVPYGSYPRILMLAWIGKIMAKKRGFGQTVEDRRISLGNLSSEFMYNIGIPMANG

gene  
complement(3120. .3323)  
/gene="csp03"  
/note="Rsp0002; RS05790"  
complement(3120. .3323)  
/gene="csp03"  
/function="miscellaneous; not classified regulator"  
/note="Product confidence : probable  
Gene name confidence : probable  
predicted by Codon\_usage  
predicted by Homology  
predicted by Framed"  
/codon\_start=1  
/evidence-not\_experimental  
/transl\_table=1  
/product="PROBABLE COLD SHOCK-LIKE TRANSCRIPTION REGULATOR  
PROTEIN"  
/protein\_id="CAD17153.1"  
/db\_xref="GI:17430469"  
/translation="MATGVKHNFTKGFRTTPDGGADLFNHPSEIQSGFKLKD  
GQKVSFEVKGQKGLQASGIKPE"  
3728. .4396  
/gene="para"  
/note="Rsp0003; RS01975"  
3728. .4396  
/gene="para"  
/function="elements of external origin; plasmid-related  
functions non-bacterial functions"  
/note="Product confidence : probable  
Gene name confidence : putative  
predicted by Codon\_usage  
predicted by Homology  
predicted by Framed"  
/codon\_start=1  
/evidence-not\_experimental  
/transl\_table=1  
/product="PROBABLE PARTITION PROTEIN"  
/protein\_id="CAD17154.1"  
/db\_xref="GI:17430470"  
/translation="MSAKIVTFVNOGCGGKTVSMHATGLGVGSKTLVDMEQ  
TATRAAQAPEEKFPASVIGLPSGAMHREVRFIADYDIIVDCPAPVSPASS  
ALISDIAIIVVPSPDILMAVAKALAOAOVNTLRVRLANMVRSLAKOA  
IEIGDGDIDILMESIGRSARFRCQAGTIVGVRARAINVDIMNVEVQLLR  
EGQA"  
4393. .5397  
/gene="parB"  
/note="Rsp0004; RS01976"  
4393. .5397  
/gene="parB"  
/function="macromolecule metabolism; macromolecule  
synthesis, modification; dna - replication, repair,  
restr./modif."  
/note="Product confidence : putative  
Gene name confidence : putative  
predicted by Codon\_usage  
predicted by Homology  
predicted by Framed"  
/codon\_start=1  
/evidence-not\_experimental  
/transl\_table=1  
/product="PUTATIVE PARTITIONING PROTEIN"  
/protein\_id="CAD17155.1"  
/db\_xref="GI:17430471"  
/translation="MSAKLSLKDMLAGIAEKTAKAPVDRFALACQATISHPRLG  
GQAPAPSAESQAESSEKRVVPIPERLOENPLNARRIYDOOIYQERASISHTGO  
QFGLAPDPAKPMYVLIIDCHYKRRALAAAGKAMECFIEDGSDIDFYLSFVINE

Query Match 5.5%; Score 101; DB 1; Length 216050;  
Best Local Similarity 48.0%; Pred. No. 2e-09;  
Matches 324; Conservative 0; Mismatches 345; Indels 6; Gaps 1;

```

gene
123. .1527
/ gene="eg1"
-35_signal
123. .128
/ gene="eg1"
-10_signal
142. .147
/ gene="eg1"
247. .1527
/ gene="eg1"
/ note="putative"
/ codon_start=1
/ transl_table=11
/ product="beta-1,4-endoglucanase"
/ protein_id="AA61980.1"
/ db_xref="GI:151208"
/ translation="MRCPMLVAASVAALMLAGCGGDDPSLSTASVSATDTTLKP
AATSTSVWLTLAKDSAAFTVSTGRVYRAGSAAWEKSVSGGCTSTFPKDPAP
GVAKVOLLGCTGCTLMRGVSLAGAFEGESLPGTGSNIITPSASVITYKKKAML
VRLPFRMERLQPTLNQFDANELSRLLGFVNATATGQVLLDPHNTARYGVNIGSS
AVPNSAYADFWRLATQFNSNPVILGLMEPSMPTEDMLSGAMLELAIRSNASN
VVFPGNATGASHMNQMTGTNGTVMKGIINDPHLVFEVHQYDGDSSGSANCV
SATIGAORLDDFTTTLRSNGYRGFLGEGFAASNDTCNOAVSNMLTFVKNNADVWTGA
MMAGPMMGGYMSIEPSNGVDKPKMSVLAPLYK"
247. .381
/ gene="eg1"
/ gene="eg1"
382. .1524
/ gene="eg1"
/ product="beta-1,4-endoglucanase"
/ note="putative"
/ citation=[2]
BASE COUNT 279 a 492 c 493 g 276 t
ORIGIN
Query Match 5.6%; Score 102.6; DB 1; Length 1540;
Best Local Similarity 48.1%; Pred. No. 1.1e-09;
Matches 325; Conservative 0; Mismatches 344; Indels 6; Gaps 1;

QY 652 CACGGTGGAGGGCAAGCTGGACAGCTCACTGGGCTCTTACACAAGAGTGTCAAGC 711
DB 762 CACGCTCAACAGAGTGTTCATGTGGAACGAGCTTCGCGCTGACCGGTTTGTCAAGC 821
QY 712 CTGCTGAGACGGGCGCTACTGATGATTTGACATGACACAACTTGGCCGCTCAACAG 771
DB 822 CGTACGCGCAGCGCCGACAGAGCGTGTCTGATCCGACACATATGCGGCTCTACGG 881
QY 772 CGGATCATCGCCAGGAGGCGGTGTGCGAGACATCTTGTGCACTCTGGTGCAGAT 831
DB 882 CAACGTATGCGGGTGAAGCGGGGCGCCACAGCGGTAGCGCATTTTCGGGGCGCT 941
QY 832 CGCAATACTACGAGACACAACAGATCATTTTGGCCCTGATGACAGCGCCAGCA 891
DB 942 GGCACCCAGTTCAGAGCAATCCCGCTCATCTCGGCTGATGACAGCGCCAAATTC 1001
QY 892 CTTGACATTTAGATCTGGGGGCGAGAGTGGCAAAAGTGTCACTGGATCGCAAGGC 951
DB 1002 GATGCCAGCAGAGAGTGGCTGTGCGGTGCCAACGCCGAGTGGCCGATCCCTCGGC 1061
QY 952 CGGCGCACCTCGCAGATGATCTCTCTGCCCGGAACCACTTTCCAGCGTGCAGAGTA 1011
DB 1062 CAATGCGAGCAAGCTGTGTTGTCGGGGGACGCGCTGGAGCGGGCGGACACTGGA 1121
QY 1012 TGTGTCTACCTGGCAGCGCGGAGCCCTGCGGACAGATTACGACCCGATGAGACCGA 1071
DB 1122 CCAGAACTGTAGGCGACGCGGAGGACGCGTATGAAAGGCGATCAATGACCGGGCCA 1181
QY 1072 TTTGCTCTTTGATTTCCACAGATCTGACATCAACAATCGGGGTGCGAGCGGA 1131
DB 1182 TTAACCTGTCTTGGAGTGCACAGTACTGATGAGCGATTCGTCGGCCAGTGGCCAA 1241
QY 1132 GTGCACACAGACAGCTGACGCC-----TTTCAACAGCTTTCGCGAGTGGCTGAGGCA 1185
DB 1242 CTCGCTGAGGCGCACATTCGGCGCCAGCGGCTGCAGGACTTTCACACCTGCGGTGGCGAG 1301

QY 1186 GAACAAGCGCCAGCCATCATCTCCGAACGCGCGCTGCATGGAACCTTGTGCATGAC 1245
DB 1302 CAACGGCTACCGCGGCTCTCGTGGGAGATTGCGGGCGCGCTGCCAACAGACACTCCAA 1361
QY 1246 TGCCTTCTGCGCCCGCAGAACAGAGCCATTAGCGAAACAGCCAGCTTACATTTGCT 1305
DB 1362 GGCCTCAGCAACATGCTCACCCTTCGTGACAGACAGCCGATGTGTGAGAGGCTGGC 1421
QY 1306 GGGCTGGGGTGCAGG 1320
DB 1422 GTGCTGGCGGGGCGG 1436

RESULT 10
AL646076 216050 bp DNA linear BCT 07-DEC-2001
LOCUS AL646076 216050 bp DNA linear BCT 07-DEC-2001
DEFINITION Ralstonia solanacearum GM11000 megaplasmid, complete sequence;
ACCESSION AL646076 AL646053
VERSION AL646076.1 GI:17430467
KEYWORDS
SOURCE Ralstonia solanacearum.
ORGANISM Ralstonia solanacearum.
Bacteria; Proteobacteria; beta subdivision; Ralstonia group;
Ralstonia.
REFERENCE 1 (bases 1 to 216050)
AUTHORS Salanoubat,M., Genin,S., Artiguenave,F., Gouzy,J., Manganot,S.,
Arlat,M., Billault,A., Brotier,P., Camus,J.C., Catolico,L.,
Chandler,M., Choise,N., Claudel-Renard,C., Cunnac,S., Demange,N.,
Gaspin,C., Lavie,M., Moisan,A., Robert,C., Saurin,W., Schlex,T.,
Siguer,P., Thebaud,P., Whalen,M., Wincker,P., Levy,M.,
Weissenbach,J. and Boucher,C.A.
Genome sequence of the plant pathogen Ralstonia solanacearum
Unpublished
2 (bases 1 to 216050)
REFERENCE Boucher,C.A.
TITLE Direct Submission
JOURNAL Submitted (05-DEC-2001) Genoscope and CNRS UMR-8030, 2 rue Gaston
Cremieux, CP5706, 91057 Evry Cedex, France, Laboratoire de Biologie
Moléculaire des Interactions Plantes-Microorganismes INRA-CNRS,
BP27, 31326 Castanet-Tolosan Cedex, France, Fondation Jean
Dausset-CEPH, 27 rue Juliette Dodu, 75010 Paris, France, LMGM CNRS
118 Route de Narbonne, F 31062 Toulouse Cedex 4, Genoscope and INRA
URGV, 2 rue Gaston Cremieux, CP5706, 91057 Evry Cedex, France,
Laboratoire de Biométrie et Intelligence Artificielle INRA, BP27,
F31326 Castanet-Tolosan Cedex, Laboratoire de Génétique Cellulaire
INRA, BP27, F31326 Castanet-Tolosan Cedex.
Christian.Boucher@toulouse.inra.fr
http://sequence.toulouse.inra.fr/R.solanacearum.html.
FEATURES
source
location/Qualifiers
1. 216050
/organism="Ralstonia solanacearum"
/strain="GM11000"
/db_xref="taxon:305"
/plasmid="megaplasmid"
1. 17
/ note="ori or RS05845
HYPOTHETICAL PARA BOX
predicted by Homology"
/ evidence=not_experimental
52. 68
/ note="RS05846
HYPOTHETICAL PARA BOX
predicted by Homology"
/ evidence=not_experimental
complement(77. 93)
/ note="RS05847
HYPOTHETICAL PARA BOX
predicted by Homology"
/ evidence=not_experimental
159. 175
/ note="RS05848
HYPOTHETICAL PARA BOX

```

```

/gene="gltb"
/notes="identified by sequence similarity; putative; ORF
located using Blastx/Glimmer/Genemark"
/codon_start=1
/transl_table=11
/product="glutamate synthase beta subunit"
/protein_id="AA034924.1"
/db_xref="GI:21106073"
/translation="MSRKAQFQFLDLBRTMPRIPIVELRTSGDMGLYGRFDKADQY
OAGRLDGNPNYCKMCPYHNAIPOMLOJVEDRIIEAALHSTPLDEYGVNCPD
DLRCSGSLTEEFAGVTGAEKYYIDTFKMGWRDLGNVAGSGRVAIVIGAPGL
ACADRLARAGITEAVYDREQITGLOEIPFSLDKSVIGRELLEBMGVQFRLG
EYGRVSTIEQLLEEDAVFLTGAVRYTDGLPGDULKVLAFLPLVNSRVISND
PHGRJAGWEDQALPDLTKRNVVVGDDTMDCVRSARIGAAVTCAYRDEASM
PSAREVANAREGEVFLFNROPISGADDEALITVETKLGEPDASGRNAVY
EGSESLLEADVIIVAFSPFLPDMLASGVAGSNGRIVAPSDMGRLPYDTTNRL
FAGGCVGRADIVYAVAGSRAGASTVGLLVKQVKKVPPAA"
complement(6763..11235)
/gene="gltb"
/notes="XAC0033"
/complement(6763..11235)
/gene="gltb"
/notes="identified by sequence similarity; putative; ORF
located using Blastx/Glimmer/Genemark"
/codon_start=1
/transl_table=11
/product="glutamate synthase alpha subunit"
/protein_id="AA034925.1"
/db_xref="GI:21106074"
/translation="MAPRTROGLNDGLYNSRDERDADCEKVAQDLDPQSRALYDT
ATAALSRTHRGVADGLTGGCGLLIKPAPFLGLARDGITLGTXYAGVYVFL
LDEADARCARAELEAOLHNAVHFRGWRVPPDDSCGLADDTLPRMOVYVDAE
QDEDFTLALFLARRAEQALHAHEFYVTLSPNISGAWLPDKLSTFYDLORS
DSSSAIVHORESTNTLPRMPLAHNFLLAHNEINTIEGRNRMAQSKWOPFRS
DIAODPVTSMHSDSOLNMLLELIAGMLOLARLVLPATPSLEFKADLAAF
VEFYGLNTEPNMGSPAGTACDSRYAACMDRNLRRARMLTSDRHEFLVASEGWEL
PAERTRKGLLPGEMMAIDLRGLDLSDAIDRLINRAAPYKQWQDQVLTQLELI
DSLVEPESEQLSRSTYKLFQLSTEVEVQVLRPLAEDEATSGMDTTPAVLSRO
TPRLVDFEFOAEAOVTNPIDLRREGIAMSITQLRETNIFHAGLEVYVNLNPNV
LSORKROLKMEQYVERNLIDLSYSEEGKAGLERICOEVEAARAGAAVLLSD
RVPDRPMALHANTGAVHHILCKVGLCDVNLITETGTAADPHHMACLLVGATAY
VYVLAYQTLFDLGRGIIQLSGGDSQIGRRYKRTYKGLSKITSKGCTTASTRG
AQLFEIVGLNDPVDLCFADTPARIGVDLARDLVEARLTVANMDQLKPEVGLK
YVHGEYHMYNDVMTLORATRTGDMQKQYVDAVHAPSTLIDLVLKRDPT
PLDDVAPADEVLRREFDSALISGALSPEHNEALIMNLRSGNSGEGEDPARYT
EKRSKIKOYASGRFQYTPXYLVNAEVLQIKVAGAKPGEGQLPGHKVNELIARLYA
KRGILISPPHHDIYSIEDLAIYDLKQVNPALVYKLVAHAGVGTIAGVYKAG
ADLTYSGHGDSGTGSPVSTIRACYPMELGVASHQALVANDLERITLQDGLKT
GLDVKKALLGADSFSGTAPMIVLCKYLRLCHLNKCTGVAADDERLRAQYFTGLP
ERVEHFFLLAEVSRQWLSYLGVRSLDEIVGRDLEQDIDIPRPGKVDLRLTHA
OYDGSACAQRLYESPDSLATQMDGLADAIANKSGDHRFLIHNTDRISGLRSGAI
ARAHNGHMSDAPLNLRFPGTGOSEFGANAGGLLEEGEANDVYKMGAGGRILVR
PPRGAERFARMNPYIVGNTCLGATGSGELFAARAGRGRVFRNSGALAVYEGSDHCE
YMTDGVVLVKGVLNFGAGFTGGLAYVLDIERDVRDYNHLEIDHRSAGCFEHR
OHLRLTISHRLETSIMVQIILDEFRDYIGKFWLKPAASTESLTERLRAA"
BASE COUNT 1900 a 3652 c 3743 g 2071 t
ORIGIN

```

Query Match 7.0%: Score 128; DB 1; Length 11366;

Best Local Similarity 51.4%: Pred. No. 1.3e-14;

Matches 377; Conservative 0; Mismatches 345; Indels 12; Gaps 3;

```

655 GGTGAGACGCAAGCTGAGAGCTCAACATGGGCTCTCTACAAAGAGTGTCAACGCGT 714
Db 2959 GCTCAATGGCCCTGGACCAAGCCCAATTTGGGGCTGATCAAGAGTCCGTCAAGCAGC 2900
715 TCTGAGAGCGGGGCTTGTGATGATGATGACACAACTTTGCCCGTACAGCGCG 774
Db 2899 CAAGGCCAACAGCAATACCTGATCTGATGTCGACAACTATGACACAGTACAGCGGCA 2840
775 CATCATCGCCGACAGGAGCGCTGTGCGACGACATCTTTGTCGACCTCTGCGTCAGATCGC 834
Db 2839 GCGCATCGGCACACGAGCGTGGCGCGCTGTGCGCGATCTGTGCGCGCGCGCTGGC 2780

```

```

835 AAAGTACTAGAGACACAGCAAGATCATCTTTGGCCCTGATGACAGCGCAGCAT 894
Db 2779 GCTGGAATTAGAGACACAGCAAGCGGTGATCTTGGGCTGATGACAGCGCAGCAT 2720
895 CGACATTGAGATCTGGCGCAGACAGCTGCCAAAGGTCGTACGTCCGAGAAAGCGCG 934
Db 2719 TTCCGCGCCCGCATGGCGCAATGCGCGGAGGAGCATGCTGCGATCCGCAAGCGCG 2660
955 CGGCACCTGCGCATGATGATCTCTGCGCGGAGCAACCACTTTGCCAGGTGAGAGTGT 1014
Db 2659 CGGCA--AATACTGATCTCTGCTGCGCGGACCGCTTACACGGCGCACAGCTGGCG 2603
1015 GTCCACTGCGAGCGCGGAGACCTTCGCGCAAGATTAGC---AACCGGATGGAAGCACCGA 1071
Db 2602 CAGACACAGCTAGCGGCTCTCCAAAGCAAGAGGGCGGTGAATCTCAAGAACCGGGTAA 2543
1072 TTGCTGTACTTTGATGTCCAAAGTATCTGACATACACACTCGGGTCCGACCGCGA 1131
Db 2542 CAACCTTGCCCTTGAAGCGCATCATACATGATGATGATGATGATGATGATGATGATGAT 2483
1132 GTGACACAGCAAGACAGCTGACGCTTCAAC-----GACTTGGCGGACTGGGTGAGGCA 1185
Db 2482 ATGACACAGCAAGATTCGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2423
1186 GAACAAACCGCCAGCGCATCATCTCCGAACGCGCGCTCCATGGAACCTTCGTGATGAC 1245
Db 2422 GAACAAACGCAAGAGGCTTCTCTGGGCGAATTCGCGACGCGCCACAAATCGGTCTGCGACA 2263
1246 TGCCTTGTGCGCCGACAAACAGGCAATGACGAAACAGGAGCTCTACATTGGCTTGT 1305
Db 2362 GGGCGTGGAAAGGATGCTGACCTACAGAAAGAAAGACGAGCTGTGGCTGGCGTGGAC 2303
1306 GGGCTGGGGTGGCGGCGAGCTTTGACAGCTGCTACATCTGACTGCTCCCTCGGCA 1365
Db 2302 CTGCTGGCGGCGCGCGCGCTGTGGAGAGCGGACTATCCTTACCGTGCACCGGGCA 2243
1366 GCCCGCACTACA 1379
Db 2242 GAGCGCGAGCGACA 2229

```

RESULT 9

PSBGL

LOCUS

DEFINITION

cds.

ACCESION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

1 (sites)

AUTHORS

1. Huang, J. Z., Sukorhman, M. and Schnell, M. A.

TITLE

1. Excretion of the egl gene product of *Pseudomonas solanacearum*

JOURNAL

1. J. Bacteriol. 171 (7), 3767-3774 (1989)

MEDLINE

89291722

PUBMED

2738021

REFERENCE

2 (bases 1 to 1540)

1. Huang, J. Z. and Schnell, M. A.

1. Role of the two-component leader sequence and mature amino acid sequences in extracellular export of endoglucanase Egl from *Pseudomonas solanacearum*

JOURNAL

1. J. Bacteriol. 174 (4), 1314-1323 (1992)

MEDLINE

92138626

PUBMED

1735723

COMMENT

1. On Feb 8, 2002 this sequence version replaced gi:342047.

1. Location/Qualifiers

1. 1540

1. /organism="Ralstonia solanacearum"

1. /strain="Am"

Db 2293 GGACCCCGGACAACTGCTGGCTTCGAAAGCGCATCACTACTGACGCGCTACACGG 2234  
 QY 1120 GTGCGACGAGTGCACACAGACAGTGCAGCCTTCAA-----CGACTGGGGA 1173  
 Db 2233 CACCAACCCGGAATGCACACAGTGCAGGCGGAGAGAGAGTGGCGGCTTTTACC 2174  
 QY 1174 CTGGCTGAGGACAGAACAGGCGCCAGGCCATCATCTCCGAAAGCGCGCTCATGAAAC 1233  
 Db 2173 CTGGCTGCGGAGAGAACAGAGAAAGGCTTCCTGGGAGAGTGGCAGCGCCACATGC 2114  
 QY 1234 TTGCTGATGACTGCTTCTTGGCCAGAACAGGCCATTAGCAAAACAGCAGCTCTA 1293  
 Db 2113 GGTCTGCAAGACGCGCTGAAAGCATGCTGACATGAGAAAGAACAGCAGCTGTG 2054  
 QY 1294 CATGGCTTTGTGGGCTGGGGTGGCGGAGCTTGACACGTCGACATCTGACTGCAC 1353  
 Db 2053 GCTGGGCTGGACCTGTTGGGGGCGCGCGCTGGTGGAGAGCGGACCTTCAATGT 1994  
 QY 1354 TCCCTGCGACAGCCCGCACTACA 1379  
 Db 1993 GCAGCGGCGCAGAGAGCGGACGAGCA 1968

RESULT 8  
 AE011626/c 11366 bp DNA linear BCT 29-MAY-2002  
 LOCUS Xanthomonas axonopodis pv. citri str. 306, section 4 of 469 of the  
 DEFINITION complete genome.  
 ACCESSION AE011626 AE008923  
 VERSION AE011626.1 GI:21106069  
 KEYWORDS  
 SOURCE Xanthomonas axonopodis pv. citri str. 306.  
 ORGANISM Xanthomonas axonopodis pv. citri str. 306.  
 Bacteria; Proteobacteria; gamma subdivision; Xanthomonas group;  
 Xanthomonas.

REFERENCE 1 (bases 1 to 11366)  
 AUTHORS da Silva,A.C.R., Ferro,J.A., Reinach,F.C., Farah,C.S., Furlan,L.R.,  
 Quaggio,R.B., Monteiro-Vitorello,C.B., Van Sluys,M.A., Almeida  
 Jr.,N.F., Alves,L.M.C., do Amaral,A.M., Bertolini,M.C.,  
 Camargo,L.E.A., Camarotte,G., Cannavan,F., Cardoso,J.,  
 Chambergo,F., Ciapina,L.P., Cicarelli,R.M.B., Coutinho,L.L.,  
 Cristiano-Santos,J.R., El-Dorry,H., Faria,J.B., Ferreira,A.J.S.,  
 Ferreira,R.C.C., Ferro,M.I.T., Formighieri,E.F., Franco,M.C.,  
 Greggio,C.C., Gruber,A., Katsuyama,A.M., Kishi,L.T., Leite  
 Jr.,R.P., Lemos,E.G.M., Lemos,M.V.F., Locali,E.C., Machado,M.A.,  
 Madeira,A.M.B.N., Martinez-Rossi,N.M., Martins,E.C., Medeiros,J.,  
 Menck,C.F.M., Miyaki,C.Y., Moon,D.H., Moreira,L.M., Novo,M.T.M.,  
 Okura,V.K., Oliveira,M.C., Oliveira,V.R., Pereira Jr.,H.A.,  
 Rossi,A., Sena,J.A.D., Silva,C., de Souza,R.F., Spinola,L.A.F.,  
 Takita,M.A., Tamura,R.E., Teixeira,E.C., Tezza,R.I.D., Trindade dos  
 Santos,M., Truffi,D., Tsai,S.M., White,F.F., Setubal,J.C. and  
 Kitajima,J.P.  
 TITLE Comparison of the genomes of two Xanthomonas pathogens with  
 differing host specificities  
 JOURNAL Nature 417 (6887), 459-463 (2002)  
 MEDLINE 22022145  
 PUBMED 12024217  
 REFERENCE 2 (bases 1 to 11366)  
 AUTHORS da Silva,A.C.R., Ferro,J.A., Reinach,F.C., Farah,C.S., Furlan,L.R.,  
 Quaggio,R.B., Monteiro-Vitorello,C.B., Van Sluys,M.A., Almeida  
 Jr.,N.F., Alves,L.M.C., do Amaral,A.M., Bertolini,M.C.,  
 Camargo,L.E.A., Camarotte,G., Cannavan,F., Cardoso,J.,  
 Chambergo,F., Ciapina,L.P., Cicarelli,R.M.B., Coutinho,L.L.,  
 Cristiano-Santos,J.R., El-Dorry,H., Faria,J.B., Ferreira,A.J.S.,  
 Ferreira,R.C.C., Ferro,M.I.T., Formighieri,E.F., Franco,M.C.,  
 Greggio,C.C., Gruber,A., Katsuyama,A.M., Kishi,L.T., Leite  
 Jr.,R.P., Lemos,E.G.M., Lemos,M.V.F., Locali,E.C., Machado,M.A.,  
 Madeira,A.M.B.N., Martinez-Rossi,N.M., Martins,E.C., Medeiros,J.,  
 Menck,C.F.M., Miyaki,C.Y., Moon,D.H., Moreira,L.M., Novo,M.T.M.,  
 Okura,V.K., Oliveira,M.C., Oliveira,V.R., Pereira Jr.,H.A.,  
 Rossi,A., Sena,J.A.D., Silva,C., de Souza,R.F., Spinola,L.A.F.,  
 Takita,M.A., Tamura,R.E., Teixeira,E.C., Tezza,R.I.D., Trindade dos

Santos,M., Truffi,D., Tsai,S.M., White,F.F., Setubal,J.C. and  
 Kitajima,J.P.  
 TITLE Direct Submission  
 JOURNAL Submitted (28-NOV-2001) Departamento de Bioluminica, Universidade de  
 Sao Paulo, Av. Prof. Lineu Prestes 748, Sao Paulo, SP 05508-900,  
 Brazil

FEATURES  
 source location/Qualifiers

1..11366  
 /organism="Xanthomonas axonopodis pv. citri str. 306"  
 /strain="306"  
 /db\_xref="taxon:190486"  
 /note="pathovar: citri"  
 complement(349..1401)  
 /gene="egl"  
 /note="XAC0029"  
 complement(349..1401)  
 /gene="egl"  
 /note="identified by sequence similarity; putative; ORF  
 located using Blastx/Glimmer/genemark"  
 /codon\_start=1  
 /transl\_table=11  
 /product="cellulase"  
 /protein\_id="AA034921.1"  
 /db\_xref="GI:21106070"  
 /translation="MKSLLTTRRLRCRALSPFLLAAMPLAQAQSAATGLKAYAGNLG  
 AEIKSOKPGVNLIDRYPADEAYAFAGKHMVRLPLMERLORAGELDPADLA  
 LIRQAVANAKAKMYIVDVHNAKYGHITGSKYDISFENLMBRLIAFSDNAV  
 IRLNMEPIYIDPQANATAQAASIDTRKGAIRNLILVRALMTGASHSYTAQGSN  
 AVAMASIRDPINRLALEAHQYLDSSGVSIGCVSAVGERLSTFEMRLKKKG  
 LAFFGNTVTCNLALNGMLGWESNSDVWMSWMAAGMAGAVPFVQPADQGRD  
 KPMQSLSPRARITD"  
 complement(2179..3252)  
 /gene="egl"  
 /note="XAC0030"  
 complement(2179..3252)  
 /gene="egl"  
 /note="identified by sequence similarity; putative; ORF  
 located using Blastx/Glimmer/genemark"  
 /codon\_start=1  
 /transl\_table=11  
 /product="cellulase"  
 /protein\_id="AA034922.1"  
 /db\_xref="GI:21106071"  
 /translation="MRTPEYFPRRLIALALTLAAGPLAQAQKEPPRRRTAGLKY  
 LDGALGILKSLSEAKAKQVLIIDHNATYSGRIGSDVPAGLADIMRLALE  
 FKDDKAVIRGLMNEPNCISAPMANAOCITAIRTKGAKNLIYVGTATGASHRS  
 TSYGVSNAKLEILKDPGNLAFBAHQYLDKDYSGTRPVCTSDSGOEKLGFTSMR  
 ENRQKGFEGFAANNPCDKALEGLTYEKNSDVWLGTWMAAGAMWKPPTVQ  
 PGRDGSDBKPMQALISXYARRAGR"  
 3560..4618  
 /gene="yank"  
 /note="XAC0031"  
 3560..4618  
 /gene="yank"  
 /note="identified by sequence similarity; putative; ORF  
 located using Blastx/Glimmer/genemark"  
 /codon\_start=1  
 /transl\_table=11  
 /product="alcohol dehydrogenase"  
 /protein\_id="AA034923.1"  
 /db\_xref="GI:21106072"  
 /translation="MSYAHAYAAQADQPLAPVFERRAPQNDVIDIAYCVCVCHSD  
 LHTAHVNMNTLYSPVGEHIVGRVTAAGVAVGPGVGLAGVGCWDSRCSAOE  
 GEBOYCEAGTGTGVNPMRGEGENTGYGSDHIVDOKYVLTTHSDNLAAVAPLCA  
 GITTISPLAHKVGPGOKGVYGLGIGHAVAIKAMGATVYVLTTSSEKRDALRL  
 GAPVAVISKDEQAQSLYNTLDFILNTVAAPHNLDPLNLRDGAIVLVGVEQSHP  
 SDVKEFYIDMADLQAA"  
 complement(5153..6634)  
 /gene="gld"  
 /note="XAC0032"  
 complement(5153..6634)

CDS

gene

gene

CDS

```

/gene="egl"
/note="XCC0028"
complement(1918..2982)
/gene="egl"
/note="identified by sequence similarity; putative; ORF
located using Blastx/Glimmer/Genemark"
/codon_start=1
/transl_table=11
/product="cellulase"
/protein_id="AA039347.1"
/db_xref="GI:21110969"
/translacion="MPSLPSTRTTHAALALALATLPLAHAKRADNARGLKYGW
NISAEFNRKRPKGLYKDYTPAADSEYFASKGMNTRLPFLERLPTLNGELDA
AOLAIKKSVAANKKOFILDLHNAYKNKRGITGEVPPALADLWRALSEKD
DKOYFGLMNPNGISSTDMMAAOAIINTIKTGANNLITVPGATYAGASMSRSTY
GYSNAKALOTIKDKPCNNLAFEAHOLDDGY SCTKPCTSATYGADKLRFPTMLENK
OKGFLGEFTGANNAYCNDALKGMLSYMEKNSDYVLTGWTWAGAMWKPDYPRVOPG
DSEKPMSTLSEHARRAGK"
/gene="yahn"
/note="XCC0029"
3296..4354
/gene="yahn"
/note="identified by sequence similarity; putative; ORF
located using Blastx/Glimmer/Genemark"
/codon_start=1
/transl_table=11
/product="alcohol dehydrogenase"
/protein_id="AA039348.1"
/db_xref="GI:21110970"
/translacion="MSKAHAYAAOTADOPAFVEERRAPGDVQIDYAGVCHSD
LHTANENHNTYPSVPGHEIVRTAVGSAVTNFKVGLAGVSCVSCRCASCOE
GEOYCEOGFTCTGNGPMGSGENTYGGSHIIVDOKVLIHSHDNLAAVAPLCA
GITTSPLAHMKVPGOGKVGVGGLGIGHMAVITAKAMGATVLFSTSSKADALRL
CASEVVISKDEQAQAQYNTLDFIINTVAAAPNIDPLNALKRDGAMVYGVPEHSH
SPAVNLVKKRTLGLSLGIGIRTOEMLDPCAKHNIVSDIETIRADQINEAYERRLK
GOVKYRFVIMDTLAKKA"
complement(5335..5841)
/gene="XCC0030"
complement(5335..5841)
/gene="XCC0030"
/note="putative; ORF located using Glimmer/Genemark"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AA039349.1"
/db_xref="GI:21110971"
/translacion="VTTLALSRSTTTTLDGRNSSVIFHNVESWVHRHMOEFERDMLN
RSIDQLVFAQANLTNAELMHIGHFDEWLETTLDYWDDELFRDQIRSTWLTATYMLNGT
SPVPIIVANSSGLMHPKATWRSNQLIEGHMRLAYLIRGMIRQNTSLRSHLVLLS
LPNSPFNY"
complement(5886..7367)
/gene="gltb"
/note="XCC0031"
complement(5886..7367)
/gene="gltb"
/note="identified by sequence similarity; putative; ORF
located using Blastx/Glimmer/Genemark"
/codon_start=1
/transl_table=11
/product="glutamate synthase, beta subunit"
/protein_id="AA039350.1"
/db_xref="GI:21110972"
/translacion="MSRKAQFQFLDPLPMPTRIVELRTSGMGLYCKPKKADAY
OSGRCLDGNPKYCKCPYHNAIPOMLIVODRRIIEAALCHSNPLPEVCGRCPQ
DRLCGSCTLEEGAVTIGAVEKYIVDAFSWGMPLDGNVPSGHRAVAVGAGPLQ
SCADRLVRAIGEAIVVDRYEDQIGLLOIPSEFKLDKSVTIGRRETLBGMQAFRIGV
ETGKDVSIQLEQLGEYDAVLTGAGRTYDGGDGLKVDLPAFLVNSRISLVGND
PHGRITAGMEDIALDNLNGRKRVVLGGDMDCVRSALIRLGAACVNCAYRDESSM
PGSAREVANAAREGVRELFENRPLSIESADDEAIGVYVERKLEPPDSCGRNAYPV
BOSSELENDVITAFGFSPTVPEWLSSGVGVEAGSNGRIVAAEGGRILPYOTSNRL
FAGDCCVRGADLVVAVAEGRDAAGSIVOLLGVKAQVEPPAA"
complement(7489..11961)

```

```

/gene="gltb"
/note="XCC0032"
complement(7489..11961)
/gene="gltb"
/note="identified by sequence similarity; putative; ORF
located using Blastx/Glimmer/Genemark"
/codon_start=1
/transl_table=11
/product="glutamate synthase, alpha subunit"
/protein_id="AA039351.1"
/db_xref="GI:21110973"
/translacion="MAPRTQGLDLYGLYNSARDERDAGFGVAQILDOPSRSLVDT
ATAIASRTHRGVNAADGLTGCGCLLRKPAFLRGLRADGILGTGVAAGVFLP
IDEAARSRSELETOLAGVQFGRMVRVPPDSGKGLADTLPRIGOVVDVAGE
ORDEFTTALFLARRABOALEVNEVNTYTPNGISYKGLVDLPDKISTEYDORS
DISSAIVTHQFSTNTPRMLHPFLILANGELITEGRRRANARSKWOPRF
DIAEEDPVISHGDSOSLDNLELILAGMDLALRILVPAOTOSLEFRADLAAP
YEEYGLNTEPMWDPAGTVACDSRYAACMLDRGLRPARMLISDRHFLVASAGVEL
PAERITRKGLDGPGEEMALDKRGDLDSDAIDRINRAAPYKWLQOQVLYLQTELI
DPSLVEEPSEQTLRSYHKLFDLSTEEVQILRPLAEQATGSGMDTPMAVLSRO
TRPLVDYFROAFOYTNPIIDPLREGIAMSITTOGKETINIFHAGAEVNVILNSPV
LSOKRLRLKMEQVYVARRRLDISYSEEGKAGLERCOEVEAARDGAMLLSD
RIPVDPDRMAHALLATGAVHHHLCKVGLRCDVNLITETGADPHHMACLFGCATAV
YPLAVQTLFELDGRGILQLSKGGSQSGRRYRGRIGKLSKIIKSGICTASYRG
AQLFEIVGLDPPVVELCTEPTPARIGVSLSDLEARELIRANNDOLKPEVGLK
YHGEAEYHNPVDVMTLORATRTGDSAPMKQYVDAVHSRPASTLRDLVOLRADPT
PLDEVAAPADVLRREDTAAISLGALESPHEALAIAMNRLGGRSNGEGEDPARYGT
EKRSKIKOYASGRPGVTPEYVLAQVLOIKVNOGAPGCGOLPHKKVNEILARBYA
KRGITGISPPPHDITYSIEDLLQYDLKOVPYALVSKYLAHAGVITAGVYKAG
ADLITVSGHDGDTGASPVSSIRYAGVPEMLGVAESHQALVANDLRITLQTDGSLK
GLDVVKAALLGADSEFGTAPMIVLGEKTLRICHLLNCAQVATODERLNAVETGLP
ERHEVEFFRLLAIEVROMLSYELVRLSDLEVTGTDLEQVSPRGVYKDLRLTHA
OYDGHSCAORSLTESPDSLATQMDGLADLAIANKGSDHRLIHNTHDSIGARLSGTI
ARAHNGHNSDAPLNLRFPGTQSGRFGANAGGILDELEGANDYVGMGMAGGRILVVR
PPRGARFERNTPITYGNTCLYATGEGELFAAARABERFVRSGLAVAYEGADNICE
YMTDGVVLVLRGLVNLFGAGTGGGLAVLYDGTVDRTNHELIDHRSARGFENHR
OHLNHLISHLRGTLSIMAOQILDFEVDYIGKFWLVKPKAAISIELTSLRRAA"

```

BASE COUNT 2136 a 3737 c 3916 g 2303 t

ORIGIN

Query Match 7.3%; Score 133.6; DB 1; Length 12092;  
 Best Local Similarity 51.6%; Pred. No. 1,1e-15;  
 Matches 385; Conservative 0; Mismatches 349; Indels 12; Gaps 3;

```

QY 643 CCGCAACACACGCGTGAGCGCAAGCGTGCAGCGCTCAACTGGGGCTCCGCAACAAGG 702
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 2710 CCGCAACACACCGCTCAATGCGCGATGCGGCGCACTGGCTGTATCAAAAATTC 2651
QY 703 CCGCAACGCGCTGTCTGAGAGCGGCGCTACTGCATGATTGACATGCACAACTTGGCCG 762
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 2650 CCGGCGCGCGCGCGCAAGCGCAAGCAAGCATGCTCGATCTGGACCTGCACAACTC 2591
QY 763 CTGCAACGCGCGCATATGCGCGAGGAGCGCTGTGCGACGACATCTTGTGCACTCTG 822
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 2590 GTGCAACGCGCAAGCGATCGCGCACGACGAGGCGCGCTGCAGCACTGCGCATCTGTG 2531
QY 823 GGTGCAAGTGCAGAACTACGAGCAAGCAAGCAAGCATCTTGGCCGTGATGACGA 882
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 2530 GCGCGCGCGCTGCGCTGAGATTCAGAGACGACAAAGCGATGTTTGGCGGTATGAACGA 2471
QY 883 GCGCGACGACCTGACATTCGATGCTGGCGGCGAGCGTGCACAAAAGCTGCTACTCGAT 942
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 2470 GCCCAATGGGATTCCTCCACGACGATGGCGCGCGCGCAAGCGCGCATCAACACAT 2411
QY 943 CCGCAAGCGCGCGCGCACTCGCATGATCTCTCTGCCGGAACCACTTTCGACGCGT 1002
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 2410 CCGCAAGCGCGCGCGCAAC---AATTTGATTCGTGGTCCGCGCACGCGCTACACCGCGC 2354
QY 1003 CGAGAGCTATGTTCCACTGCGGCGGCGGAGCGCTCGGCAAG---ATTACCAACCGGA 1059
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 2353 GCACACTGGCGGACGACGATGCGCGCTCCAAATGCCAAGCGCGCTGCAGATTCTCAA 2294
QY 1060 TGAAGCACCAGATTGCTGACTTGTATGTCACAAAGTATCTCGACATCAACAACTCCGG 1119

```

	source	1..1372	/organism="Corticium rolfsii"	
			/db_xref="taxon:39291"	
BASE COUNT	331 a	388 c	305 g	348 t
ORIGIN				
Query Match	8.2%	Score 150;	DB 6;	Length 1372;
Best Local Similarity	54.1%;	Pred. No. 7.9e-19;		
Matches 419;	Conservative	0;	Mismatches 340;	Indels 15; Gaps 5.
OY	ACGAGGGCTCAACGTCTTTGGCATATCCGATACATGCGATTGTGTCACAACACAGG	656		
Db	ACGAGGGTGCACATCTCTCCGTAATCTTGCGCTTGGCAATGATGATGCCACTCTG	435		
OY	TGGACGCAAGCTGGACGAGCTCAACTGGGGCTCTACAAAGAAGTCGTCAAGCCCTGTC	716		
Db	GTTGATGATTAAACCAACCCTTCTTCAGTCCGATACACCAACCCAGCTCAGGCTGCTC	495		
OY	TGCAACGGGGCCCTACGATGATGATGACATCACAACTTTCGCCCTACACGGCGGA	776		
Db	TGGCCACGGGTCTTACGTCATGTCGACTTGCACAACTATCTCGATGGAAGGGCAGA	555		
OY	TCATCGGCGCAGGAGCGCTGTGCGAGCATATTTTGTGCGACTTGGGCTCGATGCGAA	836		
Db	TCATTGGCGAGGGTGTGTCGCAAAAGCACAATTTGCCGTGATCTGAGACTGACTACGT	615		
OY	AGTACTACGAGACACAGACAAGATCATCTTTGGCTGATGAAGCAGCCGACACCTCG	896		
Db	CCTACTACGGCAACACCTTAAGATCATCTTTGGCTGATGAAGCAGCCCTCATGATCTCA	675		
OY	A---CATGAGATCTGGGGCGCAGAGCTGCCAAAAGTGTCTACTGCGATCGAAAGCCG	953		
Db	ATTCATATCCCCGAGTGGGGCGGACATCTTCATACGTGTCGTCACCCGTTCTGTGGCGCG	735		
OY	GGCGCACTTCGAGATGATTCCTCTGCGCCGAGAACCAACTTTCGACGCTCGAGAGTATG	1013		
Db	GCTGCA---CGAATCACTCTCTCTTACCCGGTTCTTCTGGCTGAGCGACAGCACTCC	792		
OY	TGTCACATGGCAGCGCGGAAGCCCTCGGCAAGATTACGAACCCGGATGGAACACCGATT	1073		
Db	CGACGAGGCGGAGCCTTA---CTCTGTCMAATCAGATGATCTCTGTGGGGTTACCAACA	849		
OY	TGCTCTACTTTGATGTCCACAAGTATCTGACATCAACAATCTCGGGTGGCAGCCGAGT	1133		
Db	AACCTATTTTGGATGTGACAAATCTCTGACAGGACAAACAGTAGTGACCCACTCACT	909		
OY	GCACACAGACAAGCTCGACGCGCTTCAACGAACTTGGCGGAGCGCTGAGAGCGAAN--CA	1190		
Db	GGCTTACGACAAACACTGTGTCTTCAAGACGCAAGCTGACCTGGCTTCAGAGAAATGGCA	969		
OY	AGCGCAGGCAATCATCTCCGAAAGGGGCGCTCCATGGAACCTTGTGTCATGATGCTCT	1250		
Db	ACCGTACGGGCTTGTGAGAGGAGATGTGGTGGAGTACCTGTACAGCAGATTGCGAGACAT	1029		
OY	TCTGCGCCCAAGACAGGCCATTAGCGAAACAGCGAGCTCTACAT--GGCTTGTGG	1307		
Db	ATGTGGCCCAAGAGCTTCAATTCGTTCAAAGCCAAACAATAACATGACGGCGCTTGGCA	1089		
OY	GCTGGGGTGGCGGCGAGCTTGAACAGCTGTGATCATCTTGAATCTGACTCCCTCG	1361		
Db	TCTGGGCGCGAGTGCATTTGACACGACATCATGCTTAAAGCCGTCACCCGGAAG	1143		
RESULT 7				
AE012096/c	12092 bp	DNA	linear	BCT 23-MAY-2002
LOCUS	AE012096	12092 bp	DNA	BCT 23-MAY-2002
DEFINITION	Xanthomonas campestris pv. campestris str. ATCC 33913,			
ACCESSION	AE012096 AE008922			
VERSION	AE012096.1 GI:21110967			
KEYWORDS	Xanthomonas campestris pv. campestris str. ATCC 33913.			
SOURCE	Xanthomonas campestris pv. campestris str. ATCC 33913			

Bacteria; Proteobacteria; gamma subdivision; Xanthomonas group;  
Xanthomonas.

REFERENCE AUTHORS	TITLE	
1 (bases 1 to 12092) da Silva,A.C.R., Ferro,J.A., Reinach,F.C., Farah,C.S., Furlan,L.R., Quaggio,R.B., Monteiro-Vitorello,C.B., Van Sluys,M.A., Almeida Jr.,N.F., Alves,L.M.C., do Amaral,A.M., Bertolini,M.C., Camargo,L.E.A., Camarotte,G., Cannavan,F., Cardozo,J., Chambergof,F., Ciapina,L.P., Cicarelli,R.M.B., Coutinho,L.T., Cursino-Santos,J.R., El-Dorri,H., Faria,J.B., Ferreira,A.J.S., Ferreira,R.C.C., Ferro,M.I.T., Formighieri,E.F., Franco,M.C., Greggio,C.C., Gruber,A., Katsuyama,A.M., Kishi,L.T., Leite Jr.,R.P., Lemos,E.G.M., Lemos,M.V.F., Locali,E.C., Machado,M.A., Madelara,A.M.B.N., Martinez-Rossi,N.M., Martins,E.C., Melandris, Menck,C.F.M., Miyaki,C.Y., Moon,D.H., Moreira,L.M., Novo,M.T.M., Okura,V.R., Oliveira,M.C., Oliveira,S.R., Pereira Jr.,H.A., Rossi,A., Sena,J.A.D., Silva,C., de Souza,R.F., Spínola,L.A.F., Taktla,M.A., Tamura,R.E., Teixeira,E.C., Tezza,R.L.D., Trindade dos Santos,M., Truffi,D., Tsai,S.M., White,F.F., Setubal,J.C. and Kitajima,J.P.	Comparison of the genomes of two Xanthomonas pathogens with differing host specificities <i>Nature</i> 417 (6887), 459-463 (2002)	
JOURNAL MEDLINE PUBMED REFERENCE AUTHORS	2 (bases 1 to 12092) da Silva,A.C.R., Ferro,J.A., Reinach,F.C., Farah,C.S., Furlan,L.R., Quaggio,R.B., Monteiro-Vitorello,C.B., Van Sluys,M.A., Almeida Jr.,N.F., Alves,L.M.C., do Amaral,A.M., Bertolini,M.C., Camargo,L.E.A., Camarotte,G., Cannavan,F., Cardozo,J., Chambergof,F., Ciapina,L.P., Cicarelli,R.M.B., Coutinho,L.T., Cursino-Santos,J.R., El-Dorri,H., Faria,J.B., Ferreira,A.J.S., Ferreira,R.C.C., Ferro,M.I.T., Formighieri,E.F., Franco,M.C., Greggio,C.C., Gruber,A., Katsuyama,A.M., Kishi,L.T., Leite Jr.,R.P., Lemos,E.G.M., Lemos,M.V.F., Locali,E.C., Machado,M.A., Madelara,A.M.B.N., Martinez-Rossi,N.M., Martins,E.C., Melandris, Menck,C.F.M., Miyaki,C.Y., Moon,D.H., Moreira,L.M., Novo,M.T.M., Okura,V.R., Oliveira,M.C., Oliveira,S.R., Pereira Jr.,H.A., Rossi,A., Sena,J.A.D., Silva,C., de Souza,R.F., Spínola,L.A.F., Taktla,M.A., Tamura,R.E., Teixeira,E.C., Tezza,R.L.D., Trindade dos Santos,M., Truffi,D., Tsai,S.M., White,F.F., Setubal,J.C. and Kitajima,J.P.	Direct Submission Submitted (28-NOV-2001) Departamento de Biotecnologia, Universidade de Sao Paulo, Av. Prof. Lineu Prestes 748, Sao Paulo, SP 05508-900, Brazil
FEATURES source	Location/Qualifiers	
gene	1..12092 /organism="Xanthomonas campestris pv. campestris str. ATCC 33913" /strain="ATCC_33913" /db_xref="ATCC:33913" /db_xref="taxon:190485" /note="pathovar: campestris" complement(235..1287) /gene="egl" /note="XCC0027" complement(235..1287) /gene="egl"	
CDS	/note="Identified by sequence similarity; putative; ORF located using Blastx/Glimmer/genemary" /codon_start=1 /transl_table=11 /product="cellulase" /protein_id="AA039346.1" /db_xref="GI:21110968" /translation="MSASPSMPROLALCALVMICIAATMSIAQAQSPNRLLKAGYNISSG ALEIQSKRPGLNIDYRYPPASERYFPAGKOMNVRLPIIMERQAPRAQGLDQAQLA LFOAVANAANAKANQIYLIDIVHNAYKYGOKIGSKRVYPFTDLMLRLALAFKDNNVAV IFELINPEPDYIPSESNAAAQAASIDSIRATGATULILVPGALMSGASWSTVGVSNGS AVAILANRDPLNRYAIIEVHOYLDPDSDSGTSAGCVSRITIGERLSFGSLPAQCKREFP LGEEGRANNATCTGAIDGMGTGETLNNDIVIGMTFMAGAANKKSYTFPNQPDAGGD KKPMKLUSARAHVTR"	
gene	complement(1918..2982)	

TITLE EgIII, a new endoglucanase from *Trichoderma reesei*: the characterization of both gene and enzyme

JOURNAL Gene 63 (1), 11-22 (1988)

MEDLINE 88255850

PUBMED 3384334

COMMENT Draft entry and computer-readable sequence for (1) kindly provided by M.Salohelmo, 26-MAY-1988.

FEATURES

source 1..1849

organism="Hypocrea jecorina"

/db\_xref="taxon:51453"

<1..105

/gene="EgIII"

/number=1

106..228

/gene="EgIII"

/number=1

229..590

/partial

/gene="EgIII"

/note="first expressed exon"

/number=2

join(262..590,765..1692)

/gene="EgIII"

join(262..590,765..1692)

/gene="EgIII"

/ec\_number="3.2.1.4"

/note="precursor"

/codon\_start=1

/product="endoglucanase III"

/protein\_id="AAA34213.1"

/db\_xref="GI:170549"

/translation="MNSVAPLLLAALTYGAVAAQOTVMCGGIGMSGPTNCVAGSACSLNPYACIGATITTTSTRPSPPTTTATSTSSSTPTSSSVRAGVNIAGFDRCCTDGTCTVTSKYRPLKNTFGSNRPICIGOMGFVNEDEHTIRLVGMQYLYNNIGINDSTISIKYVDLVGCSLGNVCYVDHNVARNGGIIIGCGPTNAQFTSLWSOLASKASQSRVWFQIMNEPHDVIINTMAVYQEVVTAIRMGATGQFTSLPNDMOSAGAFISDGSAAALSQVTNPDSSTNLIIPVHKYLDSDNGTHAECTNNIDGAFSPLATMIRNNROAILLETGNGVNSCIDMCQOIQYLNQNSDVIYLGIVGMGSPFSTYVLTETPTSSGSMSTDTSLVSSCLARK"

262..324

/gene="EgIII"

join(325..590,765..1689)

/gene="EgIII"

/product="endoglucanase III"

591..764

/gene="EgIII"

/number=2

765..>1692

/gene="EgIII"

/number=3

BASE COUNT 444 a 507 c 440 g 458 t

ORIGIN 367 bp upstream of Avail site.

Query Match 15.1%; Score 275.6; DB 8; Length 1849;

Best Local Similarity 59.2%; Pred. No. 4.8e-43;

Matches 489; Conservative 0; Mismatches 334; Indels 3; Gaps 1;

QY 563 GATGCGCGCGCCAGATGAGCATTTGCCGAGACGAGCGCCTCAAGCTCTTGGCAT 622

DB 832 GATGCGATCGCGCAGATGCTGCAACGAGGAGGATGACTATTTCCGCTTA 891

QY 623 TCCGCTACATGCGAGTTTCTCAACACAGCGAGCGAGAGCTGAGAGCGTCAAC 682

DB 892 CCTGTCGATGCGCATCTCTCAACACATTTGGCGGCAATCTTATTCACGAGC 951

QY 683 TGGGCGCTCTACACAGGCTGCTCAAGCGCTCTCGAGAGCGCGCTACTGATGATT 742

DB 952 ATTTCAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1011

QY 743 GACATGCAACACTTTGCCGCTACAAAGCGCGCATGCGGACGAGAGCGGTGCGAC 802

DB 1012 GACATGCAACATTTATGCTGATGAGAAAGGATGATGATGATGATGATGATGAT 1071

QY 803 GACATCTTTGTGACCTCTGGCTCCAGATCGAAAGTACTAGAGAGACAGCAAGATC 862

DB 1072 GCTCATTTACAGACGCTTTGGTGGCAGTTGGCATCAAACTACGATCTAGTCAGGGTG 1131

QY 863 ATCTTTGGCTGATGAGACGAGCGGACGACCTTCGATGATGATGATGATGATGATG 922

DB 1132 TGGTTCGATCATCATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1191

QY 923 CAAAAGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 982

DB 1192 CAAGAGTTGTAAACGCAATCCGCAACGCTGTGTACGCTCAATCTCTTTGGCT 1251

QY 983 GGAACCACTTTGCCAGCTGCGTGAAGCGATGATGATGATGATGATGATGATGATGATG 1042

DB 1252 GGAATGATTTGGCAATCTCTGCGGCTTTCATATCGATGATGATGATGATGATGATG 1311

QY 1043 AAGATTACGACACCGGATGGAAGCACCAGATTGCTGATGATGATGATGATGATGATG 1102

DB 1312 CAAGTACGACACCGGATGGAAGCACCAGATTGCTGATGATGATGATGATGATGATGATG 1371

QY 1103 GACATCAACACTCGGCTGCGACGCGGATGATGATGATGATGATGATGATGATGATGATG 1159

DB 1372 GACTCAGCAACACTCGGCTGCGACGCGGATGATGATGATGATGATGATGATGATGATGATG 1431

QY 1160 AACGACTTGGCGGACTGCGTGAAGCGACGACGACGACGACGACGACGACGACGACGACG 1219

DB 1432 TCTCGCTTGGCCACTTGGCTGCGACGACGACGACGACGACGACGACGACGACGACG 1491

QY 1220 GCGTCATGGAACCTTGTGATGATGATGATGATGATGATGATGATGATGATGATGATG 1279

DB 1492 GTTGGAACGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1551

QY 1280 AACGAGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1339

DB 1552 AACTGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1611

QY 1340 ATCTTACTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1385

DB 1612 GTCTGAGGAAACACCGACTGAGATGATGATGATGATGATGATGATGATGATGATGATG 1657

RESULT 6

E37750 1372 bp DNA linear PAT 31-JAN-2002

LOCUS E37750

DEFINITION Gene encoding endoglucanase.

ACCESSION E37750.1 GI:18624841

VERSION E37750.1

KEYWORDS JP 2000106887-A/1.

SOURCE Corticium rolfsii.

ORGANISM Corticium rolfsii

Eukaryota; Fungi; Basidiomycota; Hymenomycetes; Homobasidiomycetes; Aphyllophorales; Corticiales; Corticiaceae; Corticium.

REFERENCE 1 (bases 1 to 1372)

Yasogawa, D., Nagashima, K., Nakagawa, R. and Ikeda, T.

Gene encoding endoglucanase

Patent: JP 2000106887-A 1 18-APR-2000;

DAISUKE YASOKAWA

OS Corticium rolfsii (southern sclerotium blight) PN JP

2000106887-A/1

PD 18-APR-2000

PF 30-SEP-1998 JP 1998377864

PR

PI DAISUKE YASOGAWA, KOJI NAGASHIMA, RYOJI NAKAGAWA, TAKAYUKI IKEDA

PC C12N15/09, C12N1/15, C12N9/42//((C12N15/09, C12R1:645), (C12N9/42,

PC C12R1:69),

PC C12N15/00, (C12N15/00, C12R1:645)

CC Strandedness: Double;

CC Topology: Linear;

FH Key Location/Qualifiers

FT sig\_peptide 27..77

FT mat\_peptide 78..1196.

FT Location/Qualifiers

QY 754 CTTTCCCGCTACACGGCGGCATCAGCCAGAGGCGTGTGCGACGACATCTTTGT 813  
 Db 514 CTAGCCCTCTGGAGACGGCCAGATCATCGCAGGGTGGCCCAACAAACGACAGTTGT 573  
 QY 814 CGACCTCTGGGTCCAGATCGAAAGTACTACGAGACACAGCAAGATCATCTTTGGCT 873  
 Db 574 CTGCTCTGGAGAGAGTGGCTACAAAGTACAGAGCAACACAGAGTGGCTGGGGT 633  
 QY 874 GATACAGAGCGCCGACGAC---CTCGACATTTAGATCTGGGCGGACGACGTCGCAAAAGT 930  
 Db 634 CATATATAGAGCCCAACGACGTCGCCGACATACACAGTGGCGGACGACGTCGACAGGT 693  
 QY 931 CGTACGTCGATCCGAAAGGCGCCGACCTCCAGATGATCTCTGCGGCAACCA 990  
 Db 694 CGTACGCGCATCCGCAATCGTGGCGGTACGACGCAATAGCTGCTCGCGGGGCAATGA 753  
 QY 991 CTTTGCCAGGTCGACAGATGTGTCTCACTGGCAGCGCGGAGCCCTCGGCAAGATTAC 1050  
 Db 754 CTGACAGAGGCTGGCGCTTATGATGATGATGATGATGATGATGATGATGATGATGAT 813  
 QY 1051 GAACCCGATGACAGACCCGATTTGCTGATGATGATGATGATGATGATGATGATGATGAT 1110  
 Db 814 CAACCCGAGCGCTGACGACGACGATCTGATCTGACGTCGACAGATCTGACAGCGA 873  
 QY 1111 CACTCCGGGTGACGCGCCAGTGCACACAGACAGACGTCGAC--GCCCTCAACGACTT 1167  
 Db 874 TAACTCGGCGACGATCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 933  
 QY 1168 CGGCGACTGGTGGAGCAGAACAGCCGACGATCATCTCCGAAAGCGCGCTGCAT 1227  
 Db 934 TGCCGACTGGTGGAGCAGAACAGCCGATGATGATGATGATGATGATGATGATGATGATGAT 993  
 QY 1228 GAACTCTGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1287  
 Db 994 CACGACTCTGGAGAAATCTCTGAGACGATCTGATGATGATGATGATGATGATGATGATGATGAT 1053  
 QY 1288 CGTCTACATTTGCTTTGTGGCTGGGGTCCGCGACGCTTTGACAGCTGATCATCTTGCAC 1347  
 Db 1054 TGCTCTCTGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1113  
 QY 1348 TCTGACTCCCTGGGACGCG--CGGCACTACACGACGACGACGACGACGACGACGACGAC 1405  
 Db 1114 CCAGACGCGCATAGACGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1173  
 QY 1406 ATTCTGG 1412  
 Db 1174 ACTGTTG 1180  
 RESULT 4  
 AR030401 1155 bp DNA linear PAT 29-SEP-1999  
 LOCUS Sequence 15 from patent US 5861271.  
 DEFINITION AR030401  
 ACCESSION AR030401  
 VERSION AR030401.1 GI:5943615  
 KEYWORDS  
 SOURCE Unknown.  
 ORGANISM Unknown.  
 REFERENCE 1 (bases 1 to 1155)  
 AUTHORS Fowler,T., Clarkson,K.A., Ward,M., Collier,K.D. and Larens,E.  
 TITLE Cellulase enzymes and systems for their expressions  
 JOURNAL Patent: US 5861271-A 15 JAN-1999;  
 FEATURES  
 source 1..1155  
 1..1155 /organism="Unknown"  
 BASE COUNT 272 a 302 c 300 g 281 t  
 ORIGIN  
 Query Match 15.1%; Score 275.6; DB 6; Length 1155;  
 Best Local Similarity 59.2%; Start No. 4,9e-43;  
 Matches 489; Conservative 0; Mismatches 334; Indels 3; Gaps 1;

QY 563 GATGCGCCGCGCAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 622  
 Db 298 GATGCGATCGCGCAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 357  
 QY 623 TCCGCTACATGGCAGTTTGTCTCAACAAACGCTGACGCGGCAAGCTGACGACCTCAAC 682  
 Db 358 CTTGCTCGATGGCAGTTTGTCTCAACAAACGCTGACGCGGCAAGCTGACGACCTCAAC 417  
 QY 683 TGGGGCTCTCAACAAAGTGTGATGATGATGATGATGATGATGATGATGATGATGATGAT 742  
 Db 418 ATTCCAAAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 477  
 QY 743 GACATGACCACTTTGCGCGCTACACGCGCGCATCATGATGATGATGATGATGATGATGATGAT 802  
 Db 478 GACATCCACATTTATGCTCGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 537  
 QY 803 GACATCTTTGTCACCTCTGCGCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 862  
 Db 538 GCTCAATTCACGACCTTTGCTGCGATGATGATGATGATGATGATGATGATGATGATGATGAT 597  
 QY 863 ATCTTTGCGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 922  
 Db 598 TGTTCGCTATCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 657  
 QY 923 CAAAGGCTGCTACGCTGCGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 982  
 Db 658 CAAGAGGTTGATACCGCATTCGCAAGCTGATGATGATGATGATGATGATGATGATGATGATGAT 717  
 QY 983 GGAACCACTTTGCGACGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1042  
 Db 718 GGAATGATTTGCAATGCTGCTGGGCTTTCAATTCGATGATGATGATGATGATGATGATGAT 777  
 QY 1043 AAGATTTAGAACCCGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1102  
 Db 778 CAATGACGAACCCGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 837  
 QY 1103 GACATCAACACTCCGGCTGCGACGCGGATGATGATGATGATGATGATGATGATGATGATGAT 1159  
 Db 838 GACTCAGACACTCCGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 897  
 QY 1160 AACGACTCGCGGACTGGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1219  
 Db 898 TCTCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 957  
 QY 1220 GCTGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1279  
 Db 958 GGTGCAACGTTTCAGTCTGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1017  
 QY 1280 AACAGCGAGCTCTACATTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1339  
 Db 1018 AACTCAGATGCTATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1077  
 QY 1340 ATCTTGACTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1385  
 Db 1078 GTCTGACGGAACACGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1123  
 RESULT 5  
 TRREGIII 1849 bp DNA linear PLN 05-MAY-1994  
 LOCUS T. reesei endoglucanase III gene, complete cds.  
 DEFINITION M19373  
 ACCESSION M19373  
 VERSION M19373.1 GI:170548  
 KEYWORDS  
 SOURCE T. reesei (strain VTT-D-80133) cDNA to mRNA, and DNA.  
 ORGANISM Hypocrea jecorina  
 Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;  
 Hypocreales; Hypocreaceae; Hypocrea.  
 REFERENCE 1 (bases 1 to 1849)  
 Saloheimo,M., Lehtovaara,P., Penttila,M., Teerl,T.T., Stahlberg,J.,  
 Johansson,G., Pettersson,G., Claessens,M., Tomme,P. and  
 Knowles,J.K.



JOURNAL	Unpublished
REFERENCE	2 (bases 1 to 1373)
AUTHORS	Mernitz,G.
TITLE	Direct Submission
JOURNAL	Submitted (12-JUL-1995) G. Mernitz, Institut fuer Mikrobiologie, Max-Planck-Strasse 16, D 14512 Kleinmachnow, FRG
REFERENCE	3 (bases 1 to 1373)
AUTHORS	Mernitz,G., Koch,A., Henrissat,B. and Schulz,G.
TITLE	Endoglucanase II (EelII) of <i>Penicillium janthinellum</i> : cDNA sequence, heterologous expression and promotor analysis
JOURNAL	Curr. Genet. 29 (5), 490-495 (1996)
MEDLINE	96207475
PUBMED	8625430

FEATURES	Location/Qualifiers
source	1. .1373

```

/organism="Penicillium janthinellum"
/strain="Blouge IMET 43733 (C41)"
/db_xref="taxon:5079"
/chromosome="3"
/clone_1fb="lambda ZAPII"
1..1233
/gene="eg12"
1..1233
DDS

```

```

/gene="eq12"
/EC_number="3.2.1.4"
/codon_start=1
/product="endoglucanase 2"
/protein_id="CA61740.1"
/db_xref="GI:984166"
/db_xref="SPTRMBL:O12665"
/translation="MMADKRTLLALLALGSSALGGQDANGCGGGTACPTCVCSIG
KCSQNNMYSCLEPGASTSTSTSTTTTYYSTLLKTTTSSSAPKTCYRPAQINIG
GFDFGVVTSQTDLSQYVDESVDVENQMSHVNADTFNIFELPTGWOEIVNNNLGSGS
DSNNFGYEDQSGSLSSGAYCIVDINHAYMANGVIGQGGPTDQFISLWTLQATHK
KSNNRFIEGIMNEPHDLNTAATWATVQVTAIRNTGATSQMILLPGDYSAANEI
NGSQAALLPVTPNPDSSTTNLIEDVHKATIDNSCGTHAACVCYNNADAENLLTMYLRSNN
RQALLSTGGGNGVSCATYMKQQLIDLNANDVYLGRTSMAGGSFQASNNYVLTEVP
NGVDQIYVQCCFYRKKN"

```

Query Match	16.4%	Score	298.8	DB	8	Length	1373
Best Local Similarity	58.8%	Pred. No.	1.6e-47				
Matches	516	Conservative	0	Mismatches	362	Indels	0
						Gaps	0

Qy	477	TTCCGGGATGACATTGGCTGCGACATCGAAGCGAGCTGTCCGACTGCACAGTGGTCTG	536
Qy	478	TTTCCGGGATGACATTGGCTGCGACATCGAAGCGAGCTGTCCGACTGCACAGTGGTCTG	537
Qy	287	TTTGCGGTGTCACACTTCCCGCTTCGACTTTGGCGTGGTGCACAGTGGTACTCCAGGACC	346
Qy	537	TGCCCCCTGCTGAGCTACAAAGAGGAGATGGCGCCGCGCCAGATGAGCAATTTTCGCCGAAG	596
Db	347	TGTCTCAGAGTGGTGCATAGTCCGGGAGAGCGGTGTCAACCGATGTGCACACTTCGTCAACG	406
Qy	597	ACGAGGGGCTCAAGCTTTTGGCATATCCGATCGACATGGGAGATTTTCCCAACCAACACGG	656
Db	407	CGGATACCTTCAACACTTTTCCCGCTGCGGACTGGATGGCAGTTTCATTGTTCACAAACAACC	466
Qy	657	TGACGCGCAAGCTGGACGAGTCAACTGGGCTCCCTCAACAAGGTGCTCAACGGCCTGTC	716
Db	467	TTGGCGGTTCTCTCCAGTCGAACAACACTTTGGCAGATATGAACAACATTTGGTTCAAGGCTGT	526
Qy	717	TGCAAGCGGGGCGCTACAGCATGATGATGCATCCACAACATTTCGCCCGCTCAACAGCGCGGA	776
Db	527	CTCTCTCTGGGCGCTATTGTTATGTTGTGATTCACAACTCAACCCCGCTGAACAGGTTGGCG	586
Qy	777	TCAATCGGCTCAGGAGGCGCTGTGCGACGACATCTTTGTGCAGCTCTGGGCTCGAGATCGCA	836
Db	587	TTATCGGACAAAGCGCGTCCAAACCGATGACCAATCATCAGCCTCTGGACCGCAATTGGCCA	646
Qy	837	AGTACTACGAGACACAACAAGATCATCTTTGGCGCTGATGAACAAGCGCGCACAGACTCG	896
Db	647	CTCATCAACAAAGTACGACAGGGTATCTCTTTGGCGATATGAACAAGCGCCCATGTGATCTCA	706

QY	897	ACATTGAGATCTGGGGCGGAGACAGCTGGCAAAAGGTGTCACTGGCATTCGGAAGGCCGGCG	956
Db	707	ACATCGCAACCTGGGGCTGCCACAGTCCGGAACACTGTGATCTCGATGAGAAACACTGGAG	766
QY	957	CCACCTCGCAGATGATGTCCTCTCTGCCCCGGAACCACTTGTGCGAGCCCTCGAGAGTATGTCT	1016
Db	767	CCACGCTTCAGATGATTCCTGTCTCCCTGGTAGCGATTACACGCGCTGGCACTTCATTTG	826
QY	1017	CCACCTGGCAGCGGGGGAAGCCCTCGGCAAGATTACGACCCGGATGGAAGCACCCGATTGGC	1076
Db	827	AGAAAGGCTCGGGTGACACGCTCTCTACCGGTGACCACCCGGATGTGATCCACACGAAC	886
QY	1077	TGTACTTTGATGTGCACAGAATCTCTCGACATCAACCACTCCGGTGTGCGCCGCGAGTGA	1136
Db	887	TGATCTTTGATGTTCACAGAATCTCTCGATCTGATGATGACACAGCGGCACCCACGCCGAGTGG	946
QY	1137	CCACAGACACAGCTGGAGCGCTTCAAGCAGCTTGGCGAGCTGCTGAGGCGAAGCAACAGCCG	1196
Db	947	TGACCAACAAATGCCACGCCCTTCAACAACTTTGCAACCTGGCTGGCGCTCAAAACAGAGAC	1006
QY	1197	AGGCCATCATCTCCGAAACGGGCGGTGCATGGAACCTTCGAGCAATGACCTGCTTCTGGG	1256
Db	1007	AGGCTCTGCTTTCGAGACTGATGGCGGCAACGCTCCAAAGTTGTGCCACTTTCATGTGCC	1066
QY	1257	CCGAGAACAGGCGCCATTAGCGAAAAACAGCAGCTGTCAATTGGCTTTGTGGGCTGGGGTG	1316
Db	1067	AGCAGCTTGATATCTCTCATGTCGAATTCCTGATGTCTACCTCGGCTGAGACCTCTTGAGCG	1126
QY	1317	CCGGCAGCTTGACACGTGATCAATCTTGACCTGACT	1354
Db	1127	CTGGAGGCTTCCAGGCGCTTCTTGGAACATGTCTGACT	1164

RESULT 2	LOCUS	AB021657	1720 bp	DNA	linear	PLN 06-FEB-1999
DEFINITION	Trichoderma viride gene for endoglucanase II, complete cds.					
ACCESSION	AB021657					
VERSION	AB021657.1					
KEYWORDS	GI:4062992					
SOURCE	endoglucanase II,					
ORGANISM	Trichoderma viride (isolate:MC300-1)					
	DNA.					
	Trichoderma viride					

REFERENCE	Eukaryotes: Fungi; Ascomycota; Pezizomycotina; Sordariomycetes; Hypocreales; mitosporic Hypocreales; Trichoderma.
AUTHORS	1 (bases 1 to 1720)
TITLE	Endoglucanase II-T <sub>viride</sub>
JOURNAL	Published Only in Database (1998)
REFERENCE	2 (bases 1 to 1720)
AUTHORS	Matanabe,M.
TITLE	Direct Submission
JOURNAL	Submitted (20-DEC-1998) Manabu Matanabe, Pharmaceutical Technology Laboratories, Meiji Seika Kaisha, Ltd.; Kayama 788, Odawara-shi, Kanagawa 250-0852, Japan (E-mail: watanabumc.mesh.ne.jp, Tel:81-465-37-5106, Fax:81-465-36-2888)
FEATURES	Location/Qualifiers
source	1. .1720

```

/organism="Trichoderma viride"
/isolate="MC300-1"
/db_xref="taxon:5547"
join(171.499,683.1610)
/gene="eg12"
join(171.499,683.1610)
CDS
/gene="eg12"
/codon_start=1
/product="endoglucanase II"
/protein_id="AA36216.1"
/db_xref="GI:4062993"
/translation="MNRTPAPLLLAASILEFGAAAOQTWVGQCGGIGNSGPTSCAPGSG
ACSLINPYACIPIGRTSITTSRPPSGPTTTRATSTSSPPTSQCVRVAGVINGA
FDRCCTDGTGCVTSKSVYKPLKNFGANNYPGICOMOHFVDDMDTIFRLVGMQOYSL
NNLUGTLDSTSIKIDQLVOGCLSLVCTIIDITHNARNNGGIIGGCGPTPNAOFTS
WSOLASYSQASQRYWFGIMNEPHOVNITWMAATVOEVVTAIRNNGASQYIISPGEND

```

GenCore version 5.1.4-p5.4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

## OM nucleic - nucleic search, using sw model

Run on: May 15, 2003, 15:36:50 ; Search time 4717.51 Seconds

(without alignments)  
11264.793 Million cell updates/sec

Title: US-10-028-245-1

Perfect score: 1826

Sequence: 1 gtcgacccacgcgtcgtc.....aatcaaaaaaaaaaaaaa 1826

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : GenEmbl:\*

1: gb\_da:\*

2: gb\_hcg:\*

3: gb\_in:\*

4: gb\_om:\*

5: gb\_ov:\*

6: gb\_pat:\*

7: gb\_ph:\*

8: gb\_pl:\*

9: gb\_pr:\*

10: gb\_ro:\*

11: gb\_sts:\*

12: gb\_sy:\*

13: gb\_un:\*

14: gb\_vl:\*

15: em\_ba:\*

16: em\_fun:\*

17: em\_hum:\*

18: em\_in:\*

19: em\_mu:\*

20: em\_om:\*

21: em\_or:\*

22: em\_ov:\*

23: em\_pat:\*

24: em\_ph:\*

25: em\_pl:\*

26: em\_ro:\*

27: em\_sts:\*

28: em\_un:\*

29: em\_vl:\*

30: em\_hcg\_hum:\*

31: em\_hcg\_inv:\*

32: em\_hcg\_other:\*

33: em\_hcg\_mus:\*

34: em\_hcg\_pln:\*

35: em\_hcg\_rod:\*

36: em\_hcg\_mam:\*

37: em\_hcg\_vrt:\*

38: em\_sy:\*

39: em\_higo\_hum:\*

40: em\_higo\_mus:\*

41: em\_higo\_other:\*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	298.8	16.4	1370	8	PJEG2	X89564 P.janthinel
2	289.8	15.9	1770	8	AB021657	AB021657 Trichoder
3	286.6	15.7	1365	8	MP013914	U13914 Macrophomin
4	275.6	15.1	1155	6	AR030401	AR030401 Sequence
5	275.6	15.1	1849	8	TRRGIII	M19373 T. reesei en
6	150	8.2	1372	6	E37750	E37750 Gene encodi
7	133.6	7.3	12092	1	AE012096	AE012096 Xanthomon
8	128	7.0	11366	1	AE011626	AE011626 Xanthomon
9	102.6	5.6	1540	1	PSEGL	M84922 Ralstonia s
10	101	5.5	216050	1	AL646076	AL646076 Ralstonia
11	94.2	5.2	7218	6	I66494	I66494 Sequence 14
12	92.2	5.0	1185	8	AF440003	AF440003 Taloromyc
13	89.8	4.9	2010	1	AE233448	AE233448 Slnorhizo
14	80.4	4.4	9997	1	AE004077	AE004077 Xylella f
15	79.8	4.4	1240	8	AF054512	AF054512 Aspergill
16	79	4.3	1008	6	AX254752	AX254752 Sequence
17	78	4.3	1775	8	CPCCMC1	D1367 Cryptococcu
18	78	4.3	1775	8	S45137	S45137 CMCI-carbox
19	77.8	4.3	912	6	A69649	A69649 Sequence 2
20	77.6	4.2	1170	8	MP014948	U14948 Macrophomin
21	76.2	4.2	1303	8	AF235275	AF235275 Ralstonia
22	72.4	4.0	694	1	AF235276	AF235276 Ralstonia
23	72.4	4.0	694	1	AF235277	AF235277 Ralstonia
24	72.4	4.0	694	1	AF235278	AF235278 Ralstonia
25	72.4	4.0	694	1	AE011625	AE011625 Xanthomon
26	72	3.9	10597	1	AE235254	AE235254 Ralstonia
27	71.2	3.9	688	1	AF235250	AF235250 Ralstonia
28	70.8	3.9	688	1	AF235251	AF235251 Ralstonia
29	70.8	3.9	688	1	AF235252	AF235252 Ralstonia
30	70.8	3.9	688	1	AF235253	AF235253 Ralstonia
31	70.8	3.9	688	1	AF235255	AF235255 Ralstonia
32	70.8	3.9	688	1	AF235270	AF235270 Ralstonia
33	70.8	3.9	694	1	AF235271	AF235271 Ralstonia
34	70.8	3.9	694	1	AF235272	AF235272 Ralstonia
35	70.8	3.9	694	1	AF235273	AF235273 Ralstonia
36	69.2	3.8	694	1	AF235274	AF235274 Ralstonia
37	69.2	3.8	694	1	AF235275	AF235275 Ralstonia
38	69.2	3.8	694	1	AF235276	AF235276 Ralstonia
39	68.6	3.8	694	1	AF235264	AF235264 Ralstonia
40	68.6	3.8	694	1	AF235265	AF235265 Ralstonia
41	68.6	3.8	694	1	AF235266	AF235266 Ralstonia
42	68.6	3.8	694	1	AF235267	AF235267 Ralstonia
43	68.6	3.8	694	1	AF235268	AF235268 Ralstonia
44	68.6	3.8	720	8	CNS01AOP	AL113201 Botrytis
45	67	3.7	694	1	AF235257	AF235257 Ralstonia

## ALIGNMENTS

RESULT 1

PJEG2 1373 bp mRNA Linear PLN 13-SEP-1996

LOCUS P.janthinelium mRNA for endoglucanase2.

DEFINITION X89564

ACCESSION X89564.1 GI:984165

VERSION eg12 gene; endo-1,4-beta-glucanase; endoglucanase 2.

KEYWORDS Penicillium janthinelium.

SOURCE Penicillium janthinelium

ORGANISM Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes; Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Penicillium.

REFERENCE 1 (bases 1 to 1373) Mernitz,G., Koch,A., Hentissat,B. and Schulz,G.

AUTHORS Endoglucanase of Penicillium janthinelium-sequence, heterologous expression and promoter analysis

TITLE

**THIS PAGE BLANK (USPTO)**

Fri May 16 10:43:18 2003

us-10-028-245-3.rpr

Page 5

Db 1 MKASQILAAAAARVPAL 18

Search completed: May 9, 2003, 15:21:27  
Job time : 4.25799 secs